

Illinois Statewide Technical Reference Manual for Energy Efficiency

Version 5.0

Volume 4: Cross-Cutting Measures and Attachments

**Final
February 11th, 2016**

**Effective:
June 1st, 2016**

Table of Contents

VOLUME 1: OVERVIEW AND USER GUIDE

VOLUME 2: COMMERCIAL AND INDUSTRIAL MEASURES

VOLUME 3: RESIDENTIAL MEASURES

VOLUME 4: CROSS-CUTTING MEASURES AND ATTACHMENTS

6	Cross-Cutting Measures	6
6.1	Behavior.....	6
6.1.1	Adjustments to Behavior Savings to Account for Persistence.....	6
	Attachment A: Illinois Statewide Net-to-Gross Methodologies.....	18
1.1	Policy Context for this Information.....	18
1.2	Programs Currently Covered in this Document	19
1.3	Updating the IL-NTG Methods.....	20
1.4	Diverging from the IL-NTG Methods.....	20
1.5	Procedure for Non-Consensus Items	21
2	Attribution in Energy Efficiency Programs in General	22
2.1	Definitions.....	22
2.2	Spillover-Specific Issues	25
2.2.1	Measure Costs	25
3	Commercial, Industrial, and Public Sector Protocols	26
3.1	Core Non-Residential Protocol	28
3.1.1	Core Non-Residential Free Ridership Protocol.....	28
3.2	Core Non-Residential Spillover Protocol.....	36
3.3	Small Business Protocol	39
3.3.1	Free Ridership.....	39
3.4	C&I New Construction Protocol.....	40
3.4.1	Free Ridership.....	40
3.5	Study-Based Protocol.....	41
3.5.1	Free Ridership.....	41
3.6	Training and Technical Assistance Protocol.....	44
3.6.1	Program-Attributable Savings	44
3.6.2	Research Methods.....	44
4	Residential and Low Income Sector Protocols	46
4.1	Residential Cross-Cutting Approaches.....	48
4.1.1	Survey Design Issues.....	48
4.1.2	Participant Spillover	48
4.1.3	Nonparticipant Spillover Measured Through Trade Allies	50

4.1.4	Nonparticipant Spillover Measured from Customers	50
4.2	Appliance Recycling Protocol.....	53
4.2.1	Basic Method.....	54
4.2.2	Enhanced Method	58
4.3	Residential Upstream Lighting Protocol	59
4.3.1	Basic Method.....	60
4.3.2	Participant Spillover	63
4.3.3	Nonparticipant Spillover.....	63
4.4	Prescriptive Rebate (With No Audit) Protocol	64
4.4.1	Basic Method.....	64
4.4.2	Enhanced Method	67
4.5	Single-Family Home Energy Audit Protocol	68
4.5.1	Basic Method.....	68
4.6	Multifamily Protocol	72
4.6.1	Basic Method.....	73
4.7	Energy Saving Kits and Elementary Education Protocol	76
4.7.1	Basic Method.....	77
4.8	Residential New Construction Protocol.....	78
4.8.1	Basic Method.....	78
4.8.2	Participant Spillover	81
4.8.3	Builder Nonparticipant Spillover	82
5	Cross-Sector Protocols	84
5.1	Behavioral Protocol	84
5.1.1	Randomized Controlled Trials	84
5.1.2	Non-Randomized Designs	85
5.2	Code Compliance Protocol	87
5.2.1	Data Collection	87
5.2.2	Attribution Assessment	87
6	Appendix A: Overview of NTG Methods	89
6.1	Survey-Based Approaches	89
6.1.1	Self-Report Approach	89
6.1.2	Econometric/Revealed Preference Approach	90
6.2	Randomized Control Trials and Quasi-Experimental Designs.....	90
6.3	Deemed or Stipulated NTG Ratios	91
6.4	Common Practice Baseline Approaches	91
6.5	Market Analyses	91
6.6	Structured Expert Judgment Approaches.....	91

6.7	Program Theory-Driven Approach.....	92
6.8	Case Studies Design	92
7	Appendix B: References	94

Table of Figures

Figure 3-1. Core Free Ridership Algorithm 1	29
Figure 3-2. Core Free Ridership Algorithm 2	30
Figure 3-3. Core Free Ridership Algorithm 3	30
Figure 3-4. Small Business Free Ridership	40
Figure 3-5. Study-Based Free Ridership—Overview	42
Figure 3-6. Study-Based Free Ridership—No-Program FR Score Option #1	43
Figure 3-7. Study-Based Free Ridership—No-Program FR Score Option #2	43
Figure 4-1. NPSO Question Logic	52
Figure 4-2. Appliance Retirement Scenarios	57
Figure 4-3. Residential Upstream Lighting Free Ridership via In-Store Intercept	61
Figure 4-4. Residential Prescriptive Rebate (With No Audit) Free Ridership (Replace on Burnout)	65
Figure 4-5. Residential Prescriptive Rebate (With No Audit) Free Ridership (Early Replacement)	66
Figure 4-6. Single-Family Home Energy Audit Free Ridership—No Cost Measures	70
Figure 4-7. Single-Family Home Energy Audit Free Ridership—Discounted Measures	71
Figure 4-8. Multifamily Free Ridership for Property Managers—Non-CFL Measures	75
Figure 4-9. Multifamily Free Ridership—CFL Measures	75
Figure 4-10. Energy Saving Kits and Elementary Education Free Ridership	77
Figure 4-11. Residential New Construction Free Ridership	79

Table of Tables

Table 1-1. ICC Energy Efficiency Dockets	18
Table 2-1. Definitions	22
Table 3-1. Commercial, Industrial, and Public Sector Programs	26
Table 4-1. Residential and Low Income Programs	46
Table 4-2. Estimation of Respondents' NPSO Savings	53
Table 4-3. Calculation of Total NPSO Generated	53
Table 4-4. Net Savings Example for a Sample Population*	58
Table 4-5. Example Triangulation Weighting Approach	68
Table 4-6. IECC 2012 Building Energy Code	82

6 Cross-Cutting Measures

6.1 Behavior

6.1.1 Adjustments to Behavior Savings to Account for Persistence

DESCRIPTION

Energy efficiency program administrators are increasingly including behavior programs as part of their portfolios. These programs are characterized by various kinds of outreach, education, and customer engagement designed to motivate increases in conservation and energy management behaviors, and most commonly include participant-specific energy usage information. Savings impacts are evaluated by ex-post billing analysis comparing consumption before and after (or with and without) program intervention, and require M&V methods that include customer-specific energy usage regression analysis and randomized controlled trial experimental designs, among others (see Behavioral protocol set forth in the IL-TRM Attachment A: Illinois Statewide Net-to-Gross Methodologies for more information). As such, initial calculation of savings is treated as a custom protocol¹.

An important issue for many stakeholders is whether energy savings from behavior programs continue over time (i.e., whether they persist beyond the initial program year). Behavior programs have now been delivered for a number of years in many jurisdictions. The weight of evaluation evidence indicates that the energy-saving behaviors influenced through these programs can persist beyond the initial period of program intervention, even without continued program participation². This post-treatment savings persistence has implications for calculations of first-year savings, measure life, and cost-effectiveness testing. Accounting for persistence will yield savings and cost-effectiveness estimates that more accurately reflect the true benefits of these programs. Because annual goals are based on first-year savings, programs should only count savings attributable to first-year spending. The effect of persistence of savings beyond the first year should be included in lifetime savings calculations and cost-effectiveness testing.

The protocol below was developed to outline the adjustments that should be made to account for the persistence of savings beyond the year of program delivery. This protocol is applicable to behavior programs of any type, delivered to residential or C&I customers, that has evaluated evidence of program persistence. This general protocol should be used for any type of behavior program once supportable assumptions for persistence exist as measured by multi-year, rigorous evaluation studies.

Currently, evaluations calculate a custom value on an annual basis to estimate yearly savings. Evaluators typically use a regression analysis to estimate program effects. These regression analyses provide what is called an average treatment effect on the treated (ATT) estimate of program savings. The ATT approach takes advantage of the presence of a randomly assigned control group for each cohort that received reports in the service territory. These regressions use various methods to account for household-specific usage patterns³. Because of the experimental design, we can assume that the treatment and control groups experienced similar historical, political, economic,

¹ The protocol outlined here assumes that adjustments to remove the effects of savings from program lift (participation in other utility programs), including legacy uplift, to account for move-outs and opt-outs, to normalize for effects of weather, and any other appropriate adjustments, have been made as part of the custom calculation of savings – this final savings value is referred to as “Measured Savings” in the calculations below.

² Long-Run Savings and Cost-Effectiveness of Home Energy Reports Programs, Cadmus, October 2014. Also see additional sources in the REFERENCE TABLE below.

³ For example, a linear fixed-effects regression (LFER) model includes a household-specific intercept to account for time-invariant, household-level factors affecting energy use, and a post program regression (PPR) model uses energy use lags to account for household-specific usage in the year prior to the program.

and other events that had comparable effects on their energy use. Moreover, because these groups experienced generally similar weather conditions, it is not necessary to measure or include weather in the model specification.

However, in the case of comparing and summing savings year over year, exogenous factors, such as weather, make annual estimates non-equivalent. In particular, weather can play an important role in driving behavioral effects.⁴ As such, for this framework, evaluators will adjust for effects related to weather as part of the custom inputs to this protocol. This adjustment will provide savings using a model specification that incorporates standard weather year inputs (HDD and CDD), to be used in the calculation of annual savings as well as inputs for cost effectiveness as outlined below. This input will approximate average savings for a given weather year based upon historical data.⁵ This approach is consistent with how performance goals are developed, as well as other inputs for cost-effectiveness testing.

The protocol will become effective for residential Home Energy Reports (HERs)-type programs⁶ as of June 1, 2017 (program year 2018) - it is provided here for program planning purposes. All ongoing programs will undergo a “reset” upon institution of this protocol⁷. Regardless of any previous history of behavior program delivery, the program year ending May 31, 2018 will be assumed to be Year 1 for all HERs-type programs underway at that time for the purpose of the incorporation of multiyear measure life/savings persistence into cost-effectiveness calculations and for the application of the adjustments to annual savings as outlined below. Should any additional new programs (referred to as “waves” in the calculations below) be established in 2018 or in subsequent years, their first year will be assumed to be Year 1 for that wave – that is, each wave is tracked separately and savings are calculated separately using the approach outlined here. All residential HERs-type programs prior to June 1, 2017 will assume a 1-year measure life; the assumptions and protocols outlined below will not be applied retrospectively to any utility programs. All other types of behavior programs will continue to use a 1-year measure life until supportable evidence exists for savings persistence, at which time this adjustment protocol can be used.

DETERMINATION OF EFFICIENT BEHAVIOR

Behavior programs focus primarily on reducing electricity and natural gas consumption through behavioral changes; this reduction is generally measured through ex-post billing analysis after program intervention. Specific energy conservation and management behaviors are not usually directly observable. The specific definition of the efficient case is part of the design of behavioral programs and is included as part of the custom saving protocol, which will include any adjustment necessary to remove effects of program-related investments in efficient equipment.

DETERMINATION OF BASELINE BEHAVIOR

The ideal baseline for behavior programs is the energy usage without the program intervention. Various types of experimental, quasi-experimental, and/or regression-based EM&V approaches are used to present statistically

⁴ We acknowledge that this approach is a proxy for estimating actual savings to allow for prospective calculation of lifetime savings. However, a substantial limitation to this approach is the issue of unobserved behavioral ramp-up that is likely to occur for future waves of participants.

⁵ In the future, this approach could be empirically tested by comparing actual savings calculated in future program years against standard weather year results, producing a ‘realization rate’ between planned and actual savings results. Standard weather years could potentially be enhanced to better reflect these differences.

⁶ Residential HERs-type programs: programs that deliver Home Energy Reports to homeowners using a random control trial (RCT) experimental design. Behavior change is motivated by customer-specific usage information with individualized analysis, comparisons, and tips for energy savings.

⁷ It is understood that this approach does not accurately take into account that programs have been in place prior to this date, and the fact that customers at that time will have been receiving reports for variable amounts of time, with varied associated actual savings persistence from these earlier program efforts. The difficulties of trying to “phase in” persistence adjustments to reflect this history have been recognized, and the approach outlined here has been recommended by the Illinois TAC members as a reasonable approximation.

valid approximations to this without-program baseline⁸. The specific definition of the baseline case is part of the design of behavioral programs and is included as part of the custom saving protocol.

DEEMED LIFETIME/PERSISTENCE OF SAVINGS

Evaluations in Illinois have shown that savings from residential HERs-type behavior programs can persist into the year following program delivery⁹, though savings levels decay in the second year. For other residential RCT programs evaluated to date, savings have been shown to persist for at least 3 years year following program delivery¹⁰, and industry expectations are that savings likely persist beyond that. We assume here that savings persist at some level for 5 years¹¹. Savings over those 5 years are not equal, however; it is preferable that actual levels of ongoing savings should be calculated by future year as outlined below (see Application of Persistence for Cost-effectiveness) and used in cost-effectiveness and lifetime savings calculations¹². Measure life is assumed = 1 year for other behavior program types.

DEEMED MEASURE COST

It is assumed that most behavior changes in residential settings can be accomplished with homeowner labor only and without investment in new equipment; therefore, without evidence to the contrary, measure costs in such residential programs focused on motivating changes in customer behavior may be defined as \$0¹³. Costs for C&I programs may include additional staffing, software purchases, etc. Cost for such programs is therefore program specific and is determined on a custom basis.

LOADSHAPE AND COINCIDENCE FACTOR

While there is evidence from analysis of AMI data that the savings loadshape for residential HERs-type programs mirrors the whole-house electric energy load pattern, there are not yet enough data to develop a behavior-specific loadshape. Indications from several unpublished analyses¹⁴ show that these behavior savings occur in a general

⁸ See the Illinois Behavioral protocol set forth in the IL-TRM Attachment A: IL-NTG Methods for more information concerning randomized control trials and quasi-experimental evaluation methods for non-randomized designs for behavior programs.

⁹ ComEd Home Energy Report Opower Program Decay Rate and Persistence Study DRAFT. Energy Efficiency/Demand Response Plan: Plan Year 7 (6/1/2014-5/31/2015). Navigant, Presented to Commonwealth Edison Company, January 29, 2016; Behavioral Energy Savings Programs: Home Energy Reports Persistence Study Part 2 – April 2015 to September 2015 Nicor Behavioral Energy Savings Programs, Navigant, January 2016.

¹⁰ Long-Run Savings and Cost-Effectiveness of Home Energy Reports Programs, Cadmus, October 2014. Also see additional sources in the REFERENCE TABLE below. Given the limited persistence studies available, we acknowledge that using an average of these studies by fuel type may be the best approximation of persistence rates. However, moving forward, the TAC will incorporate additional study values and develop the most appropriate persistence factor taking into account participant characteristics, such as the duration of exposure, the frequency of reports, baseline usage, as well as the amount of time that has persisted since receiving their final report

¹¹ Determined as a reasonable preliminary assumption by Illinois TAC members. This assumption should be updated as additional research is conducted on these types of programs, and additional evaluation should be undertaken to assess the reasonableness of this assumption for Illinois-specific programs.

¹² This method of applying calculated values for future year benefits is preferred. Alternatively, an effective measure life can be calculated as Effective Measure Life = Total Discounted Lifetime Savings / First Year Savings.

¹³ Future evaluation of costs of behavior change is encouraged to help clarify this assumption. In addition, as noted earlier in this measure characterization, in order to ensure double counting of savings does not occur, the protocol outlined here assumes that adjustments to remove the effects of program lift have been made as part of the custom calculation of savings. In a similar manner, given the savings accounted for by other utility programs are removed from the savings claims and cost-effectiveness for the behavior program, the incremental costs associated with such utility program incentivized measures should also be excluded from the behavior program cost-effectiveness analysis, so as to help ensure double counting of costs does not occur in the utility portfolio cost-effectiveness analysis.

¹⁴ Based on communication from Mathias Bell based on (currently unpublished) studies done by Opower, Cadmus, and LBNL.

pattern most closely approximated by the Residential Electric Heating and Cooling Loadshape (R10) than any other current residential measure loadshape; this is therefore recommended as the most reasonable approximation for use until more-specific data are available. Loadshapes and coincidence factors will need to be determined for other types of behavior programs once sufficient data are in hand.

Algorithm

CALCULATION OF SAVINGS

Throughout these protocols, Year T refers to the current reporting year for which annual savings are being determined¹⁵.

ELECTRIC ENERGY SAVINGS

The algorithm shown below for this measure was developed to calculate the annual persistence-adjusted electric savings in to be reported in year T after adjustment to account for the proportion of the measured savings for that program year that actually reflects any persistent savings from prior years' program activities (Years T-1, T-2, T-3, and T-4)¹⁶.

$$\Delta kWh_{T \text{ Adjusted}} = \Delta kWh_{T \text{ Measured}} - (\Delta kWh_{T-1 \text{ Adjusted}} * RR_{T-1,T} * PFE_1) - (\Delta kWh_{T-2 \text{ Adjusted}} * RR_{T-2,T} * PFE_2) - (\Delta kWh_{T-3 \text{ Adjusted}} * RR_{T-3,T} * PFE_3) - (\Delta kWh_{T-4 \text{ Adjusted}} * RR_{T-4,T} * PFE_4)$$

Where:

$\Delta kWh_{x \text{ Adjusted}}$ = total program annual savings for year X after adjustments to account for persistence (calculated value)

$\Delta kWh_{x \text{ Measured}}$ = measured kWh savings: total program savings as determined from custom calculation/billing analysis¹⁷ of participants in program during year X (input value)

$RR_{Y,X}$ = Program retention rate in year X from year Y participation
= % of program participants in year Y that are still in program in year X (input value: calculated as # participants still in program in year X / # participants in year Y)

PFE_Z = Persistence factor - electric (deemed value)
= % savings that persist Z years after savings were initially measured, where Z is a number from 1 - 4

Also see DTE Energy: Behavior Program Measures for Submission to 2015 MEMD - Year Three Energy Savings - Demand Savings. Energy Optimization, April 15, 2014. http://www.michigan.gov/documents/mpsc/memd_2015_453673_7.pdf

¹⁵ Calculation algorithms account for program attrition as well as persistence decay. It has been noted that there may also be a need to adjust for cross-year effects of large differences in weather conditions or economic impacts. Custom savings inputs therefore are adjusted for standard year weather. Further studies are needed to help determine the magnitude of such effects and if this is the appropriate way to account for them.

¹⁶ This calculation should be carried out separately for each "wave" of behavior programs, where a wave is defined as a newly launched program. For simplicity, any new wave is assumed to start at the beginning of a program year (Year 1) and may include multiple different treatment types such as usage groups, report frequency, etc. For example, any wave added after 2018, will be considered Year 1 in the year they are launched.

¹⁷ All appropriate adjustments to remove effects of participation in other utility programs, move-outs, opt-outs, to normalize for effects related to weather, and other adjustments as determined by the program experimental design, are assumed to have been made to result in this value for "measured savings". This value has been adjusted for year X weather terms.

= use table below to select the appropriate value

Electric Persistence Factors¹⁸

Program Type	Program Year T - record 100% of adjusted savings ($\Delta kWh_{TAdjusted}$ above)	Percent adjusted savings from Year T activities that persist 1 year after year T	Percent adjusted savings from Year T activities that persist 2 years after year T	Percent adjusted savings from Year T activities that persist 3 years after year T	Percent adjusted savings from Year T activities that persist 4 years after year T
		PFE ₁	PFE ₂	PFE ₃	PFE ₄
Residential HERs-type (RCT)	100%	82%	68%	56%	46%

¹⁸ See REFERENCE TABLE below for sources.

Example of Adjusted Annual Savings Calculations:

Assume the following information on participation and measured savings for the following program years (all adjustments have been made to remove effects of program lift, weather, etc. within the custom savings calculations). Assume 2018 is the first year of all programs (or is the “reset” year).

	Reporting Year					
	2018	2019	2020	2021	2022	2023
Input data from program information and custom savings analysis						
# Participants (households)	120,000	109,000	103,000	99,000	94,000	90,000
kWh per participant (household)	200	250	245	250	250	265
Measured kWh savings (custom)	24,000,000	27,250,000	25,235,000	24,750,000	23,500,000	23,850,000

Calculation of Retention Rates:

For use in 2019:

$$RR_{2018, 2019} = 109,000/120,000 = 0.908$$

For use in 2020:

$$RR_{2018, 2020} = 103,000/120,000 = 0.858$$

$$RR_{2019, 2020} = 103,000/109,000 = 0.945$$

For use in 2021:

$$RR_{2018, 2021} = 99,000/120,000 = 0.825$$

$$RR_{2019, 2021} = 99,000/109,000 = 0.908$$

$$RR_{2020, 2021} = 99,000/103,000 = 0.961$$

For use in 2022:

$$RR_{2018, 2022} = 94,000/120,000 = 0.783$$

$$RR_{2019, 2022} = 94,000/109,000 = 0.862$$

$$RR_{2020, 2022} = 94,000/103,000 = 0.913$$

$$RR_{2021, 2022} = 94,000/99,000 = 0.949$$

For use in 2023:

$$RR_{2019, 2023} = 90,000/109,000 = 0.826$$

$$RR_{2020, 2023} = 90,000/103,000 = 0.874$$

$$RR_{2021, 2023} = 90,000/99,000 = 0.909$$

$$RR_{2022, 2023} = 90,000/94,000 = 0.957$$

Calculation of Adjusted Annual Savings:

$$\Delta kWh_{2018 \text{ Adjusted}} = 24,000,000 \text{ kWh}$$

$$\begin{aligned} \Delta kWh_{2019 \text{ Adjusted}} &= 27,250,000 - (24,000,000 * 0.908 * 0.82) \\ &= 9,380,560 \text{ kWh} \end{aligned}$$

$$\begin{aligned} \Delta kWh_{2020 \text{ Adjusted}} &= 25,235,000 - (9,380,560 * 0.945 * 0.82) - (24,000,000 * 0.858 * 0.68) \\ &= 3,963,444 \text{ kWh} \end{aligned}$$

$$\begin{aligned} \Delta kWh_{2021 \text{ Adjusted}} &= 24,750,000 - (3,963,444 * 0.961 * 0.82) - (9,380,560 * 0.908 * 0.68) - (24,000,000 * 0.825 * 0.56) \\ &= 4,746,794 \text{ kWh} \end{aligned}$$

$$\begin{aligned} \Delta kWh_{2022 \text{ Adjusted}} &= 23,500,000 - (4,746,794 * 0.949 * 0.82) - (3,963,444 * 0.913 * 0.68) - (9,380,560 * 0.862 * 0.56) \\ &\quad - (24,000,000 * 0.783 * 0.46) \\ &= 4,172,971 \text{ kWh} \end{aligned}$$

$$\begin{aligned} \Delta kWh_{2023 \text{ Adjusted}} &= 23,850,000 - (4,172,971 * 0.957 * 0.82) - (4,746,794 * 0.909 * 0.68) - (3,963,444 * 0.874 * 0.56) \\ &\quad - (9,380,560 * 0.826 * 0.46) \\ &= 12,137,109 \text{ kWh} \end{aligned}$$

Apply the same approach to calculate adjusted annual kW and Therms.

SUMMER COINCIDENT PEAK DEMAND SAVINGS

Coincident peak demand savings in year T should also be adjusted to account for persistence from previous years using a similar algorithm¹⁹.

¹⁹ While there are no current studies that evaluate the persistence of peak savings, without more-specific information on the

If peak demand is measured directly by the custom savings analysis:

$$\Delta kW_{T \text{ Adjusted}} = \Delta kW_{T \text{ Measured}} - (\Delta kW_{T-1 \text{ Adjusted}} * RR_{T-1,T} * PFE_1) - (\Delta kW_{T-2 \text{ Adjusted}} * RR_{T-2,T} * PFE_2) \\ - (\Delta kW_{T-3 \text{ Adjusted}} * RR_{T-3,T} * PFE_3) - (\Delta kW_{T-4 \text{ Adjusted}} * RR_{T-4,T} * PFE_4)$$

Where:

$\Delta kW_{X \text{ Adjusted}}$ = total program demand savings for year X after adjustments to account for persistence (calculated value)

$\Delta kW_{X \text{ Measured}}$ = total program demand savings as determined from custom calculation /billing analysis²⁰ of participants in program during year X (input value)

Other variables as defined above

If peak demand is not measured directly by the custom savings analysis, peak demand should be calculated as follows:

$$\Delta kW_{T \text{ Adjusted}} = (\Delta kWh_{T \text{ Adjusted Summer}} / \# \text{summer hours}) * \text{peak adjustment factor}$$

Where:

$\Delta kWh_{T \text{ Adjusted Summer}}$ = average adjusted electric energy savings (calculated above) for peak summer months

$$= \Delta kWh_{T \text{ Adjusted}} * 0.42 * (3/5)$$

$$= \Delta kWh_{T \text{ Adjusted}} * 0.25$$

Where:

0.42 = Summer Loadshape % for May – Sept

3/5 = proportion of May-Sept hours that fall in June, July, and Aug

summer hours = # hours in June, July, and Aug

$$= 8760 / 4$$

Where: 8760 = Hours per year

peak adjustment factor = adjustment for peak k/w over average kW for these hours

$$= 1.5^{21}$$

actual behaviors undertaken by program participants and their corresponding peak savings, it seems reasonable to assume that peak savings will also persist in a similar pattern; both of the approaches given assume persistence in peak savings. Further evaluation should be undertaken to clarify this point and determine appropriate peak-specific persistence values.

²⁰ All appropriate adjustments to remove effects of participation in other utility programs, move-outs, opt-outs, to normalize for effects related to weather, and other adjustments as determined by the program experimental design, are assumed to have been made to result in this value for “measured savings”. This value has been adjusted for year X weather terms.

²¹ Based on an approach used in Michigan that gives resulting values supported by evaluation claims. Also see DTE Energy: Behavior Program Measures for Submission to 2015 MEMD - Year Three Energy Savings - Demand Savings. Energy Optimization, April 15, 2014. http://www.michigan.gov/documents/mpsc/memd_2015_453673_7.pdf

NATURAL GAS ENERGY SAVINGS

The algorithm shown below for this measure was developed to calculate the annual persistence-adjusted Therm savings in to be reported in year T after adjustment to account for the proportion of the measured savings for that program year that actually reflects any persistent savings from prior years' program activities (Years T-1, T-2, T-3, and T-4).²²

$$\Delta\text{Therms}_{T \text{ Adjusted}} = \Delta\text{Therms}_{T \text{ Measured}} - (\Delta\text{Therms}_{T-1 \text{ Adjusted}} * \text{RR}_{T-1,T} * \text{PFG}_1) - (\Delta\text{Therms}_{T-2 \text{ Adjusted}} * \text{RR}_{T-2,T} * \text{PFG}_2) - (\Delta\text{Therms}_{T-3 \text{ Adjusted}} * \text{RR}_{T-3,T} * \text{PFG}_3) - (\Delta\text{Therms}_{T-4 \text{ Adjusted}} * \text{RR}_{T-4,T} * \text{PFG}_4)$$

Where:

$\Delta\text{Therms}_{X \text{ Adjusted}}$ = total program annual savings for year X after adjustments to account for persistence (calculated value)

$\Delta\text{Therms}_{X \text{ Measured}}$ = total program savings as determined from custom calculation/billing analysis²³ of participants in program during year X (input value)

PFG_Z = Persistence factor - gas (deemed value)

= % savings that persist Z years after savings were initially measured, where Z is a number from 1 - 4

= use table below to select the appropriate value

Other variables as defined above

Gas Persistence Factors²⁴

Program Type	Program Year T - record 100% of calculated savings ($\Delta\text{Therms}_{T \text{ Adjusted}}$ above)	Percent adjusted savings from Year T activities that persist 1 year after year T	Percent adjusted savings from Year T activities that persist 2 years after year T	Percent adjusted savings from Year T activities that persist 3 years after year T	Percent adjusted savings from Year T activities that persist 4 years after year T
		PFG_1	PFG_2	PFG_3	PFG_4
Residential HERs-type (RCT)	100%	45%	20%	9%	4%

APPLICATION OF PERSISTENCE FOR COST-EFFECTIVENESS

For determination of cost effectiveness (or lifetime savings) of programs in year T, future years' savings related to the current year activities should be recorded for this measure as savings for each specific year using the table

²² This calculation should be carried out separately for each "wave" of behavior programs, where a wave is defined as a newly launched program. For simplicity, any new wave is assumed to start at the beginning of a program year (Year 1) and may include multiple different treatment types such as usage groups, report frequency, etc.

²³ All appropriate adjustments to remove effects of participation in other utility programs, move-outs, opt-outs, to normalize for effects related to weather, and other adjustments as determined by the program experimental design, are assumed to have been made to result in this value for "measured savings". This value has been adjusted for year X weather terms.

²⁴ See REFERENCE TABLE below for sources.

below²⁵. Because of the potentially confounding effects of differences in weather in future years, the savings inputs used ($\Delta kWh_{TAdjusted}$, $\Delta kW_{TAdjusted}$, $\Delta Therms_{TAdjusted}$) for these future-year savings calculations have been developed using inputs that are provided from a custom model specification that incorporates standard weather year inputs (HDD and CDD). This input (to be provided by program evaluators) will approximate average savings for a given weather year based upon historical data.²⁶

Program Year T - record 100% of adjusted annual savings as calculated above	Percent savings from Year T activities that persist 1 year after year T	Percent savings from Year T activities that persist 2 years after year T	Percent savings from Year T activities that persist 3 years after year T	Percent savings from Year T activities that persist 4 years after year T
$\Delta kWh_{TAdjusted}$ $\Delta kW_{TAdjusted}$ $\Delta Therms_{TAdjusted}$	$\Delta kWh_{TAdjusted} * PFE_1$ $\Delta kW_{TAdjusted} * PFE_1$ $\Delta Therms_{TAdjusted} * PFG_1$	$\Delta kWh_{TAdjusted} * PFE_2$ $\Delta kW_{TAdjusted} * PFE_2$ $\Delta Therms_{TAdjusted} * PFG_2$	$\Delta kWh_{TAdjusted} * PFE_3$ $\Delta kW_{TAdjusted} * PFE_3$ $\Delta Therms_{TAdjusted} * PFG_3$	$\Delta kWh_{TAdjusted} * PFE_4$ $\Delta kW_{TAdjusted} * PFE_4$ $\Delta Therms_{TAdjusted} * PFG_4$

²⁵ These cost-effectiveness calculations assume a retention rate of 100% after the first program year. Move-out rates and other attrition factors continue to occur and fluctuate year over year, and, to be accurate, the value of this persistence for lifetime cost and cost-effectiveness calculations should adjust for this attrition through the application of a deemed estimate. At this time, we do not have sufficient data for such an adjustment and recommend further evaluation to develop appropriate values.

²⁶ In the future, this approach could be empirically tested by comparing actual savings calculated in future program years against standard weather year results, producing a 'realization rate' between planned and actual savings results. Standard weather years could potentially be enhanced to better reflect these differences.

Example of Calculation of Cost-effectiveness Inputs – for Electric Savings:

Assume the same information as was used in the Example of Adjusted Annual Savings Calculations (first row below). However, within each year of calculation, evaluators will adjust custom evaluation input (e.g., $\Delta kWh_{2018WeatherAdj}$) by modeling estimated savings using the standard weather year for prospective application. Assume these custom inputs provide the values in the second row below.

	Reporting Year T					
	2018	2019	2020	2021	2022	2023
Adj. kWh savings (previously calculated) $\Delta kWh_{TAdjusted}$	24,000,000	10,252,240	5,116,954	5,731,923	5,024,404	11,556,551

In 2018:

2018 annual savings = $\Delta kWh_{2018 Adjusted}$ = 24,000,000 kWh

Cost-effectiveness benefit in 2019 = $\Delta kWh_{2018 Adjusted} * PFE_1 = 24,000,000 * 0.82 = 19,680,000$ kWh

Cost-effectiveness benefit in 2020 = $\Delta kWh_{2018 Adjusted} * PFE_2 = 24,000,000 * 0.68 = 16,320,000$ kWh

Cost-effectiveness benefit in 2021 = $\Delta kWh_{2018 Adjusted} * PFE_3 = 24,000,000 * 0.56 = 13,440,000$ kWh

Cost-effectiveness benefit in 2022 = $\Delta kWh_{2018 Adjusted} * PFE_4 = 24,000,000 * 0.46 = 11,040,000$ kWh

In 2019:

2019 annual savings = $\Delta kWh_{2019 Adjusted}$ = 10,252,240 kWh

Cost-effectiveness benefit in 2020 = $\Delta kWh_{2019 Adjusted} * PFE_1 = 10,252,240 * 0.82 = 8,406,837$ kWh

Cost-effectiveness benefit in 2021 = $\Delta kWh_{2019 Adjusted} * PFE_2 = 10,252,240 * 0.68 = 6,971,523$ kWh

Cost-effectiveness benefit in 2022 = $\Delta kWh_{2019 Adjusted} * PFE_3 = 10,252,240 * 0.56 = 5,741,254$ kWh

Cost-effectiveness benefit in 2023 = $\Delta kWh_{2019 Adjusted} * PFE_4 = 10,252,240 * 0.46 = 4,716,030$ kWh

In 2020:

2020 annual savings = $\Delta kWh_{2020 Adjusted}$ = 5,116,954 kWh

Cost-effectiveness benefit in 2021 = $\Delta kWh_{2020 Adjusted} * PFE_1 = 5,116,954 * 0.82 = 4,195,902$ kWh

Cost-effectiveness benefit in 2022 = $\Delta kWh_{2020 Adjusted} * PFE_2 = 5,116,954 * 0.68 = 3,479,529$ kWh

Cost-effectiveness benefit in 2023 = $\Delta kWh_{2020 Adjusted} * PFE_3 = 5,116,954 * 0.56 = 2,865,494$ kWh

Cost-effectiveness benefit in 2024 = $\Delta kWh_{2020 Adjusted} * PFE_4 = 5,116,954 * 0.46 = 2,353,799$ kWh

Etc.

Apply the same approach to calculate cost-effectiveness inputs for kW and for Therms.

WATER IMPACT DESCRIPTIONS AND CALCULATION

N/A

DEEMED O&M COST ADJUSTMENT CALCULATION

N/A

Illinois Statewide Technical Reference Manual – 6.1.1 Adjustments to Behavior Savings to Account for Persistence

REFERENCE TABLE

Persistence: Reference Studies ²⁷							
Utility/Location	Frequency of Reports when in program	Number of Months in Program Before Terminated	Number of Post-Treatment Savings Analysis Months	Average Annual savings decay	Persistence (= 100% - decay)	Source	Electric or Gas
Upper Midwest	Monthly & quarterly	24-25	26	21%	79%	1	Electric
West Coast	Monthly & quarterly	24	29	18%	82%	1	Electric
West Coast	Monthly & quarterly	25-28	34	15%	85%	1	Electric
SMUD	Monthly & quarterly	27	12	32%	68%	1	Electric
Puget Sound Energy	Monthly & quarterly	24	36	11%	89%	1	Electric
MASS	Monthly & quarterly	26	15	33%	67%	2	Electric
Illinois (ComEd)	Bimonthly	52	12	4%	96%	3	Electric
Illinois (ComEd)	Bimonthly	30	12	2%	98%	3	Electric
Illinois (ComEd)	Bimonthly	16	12	22%	78%	3	Electric
Average Annual Electric Savings Persistence:					82%		
MASS	Monthly & quarterly	15	17	64%	36%	2	Gas
Illinois (Nicor)	Bimonthly	12	12	46%	54%	4	Gas
Average Annual Gas Savings Persistence:					45%		

Sources:

- 1: http://www.cadmusgroup.com/wp-content/uploads/2014/11/Cadmus_Home_Energy_Reports_Winter2014.pdf
- 2: <http://ma-eeac.org/wordpress/wp-content/uploads/Home-Energy-Report-Savings-Decay-Analysis-Final-Report1.pdf>
- 3: http://ilsagfiles.org/SAG_files/Technical_Reference_Manual/Version_5/Sources/Nicor_Gas_HER_Persistence_Study_Part_2_DRAFT_2016-01-28.pdf
- 4: http://ilsagfiles.org/SAG_files/Technical_Reference_Manual/Version_5/Sources/ComEd_HER_Opower_Persistence_and_Decay_Study_DRAFT_2016-01-28.pdf

MEASURE CODE: CC-BEH-BEHP-V01-170601

²⁷ These persistence studies done to date capture effects only through a limited time frame and only for the specific program characteristics of the study programs. They may not accurately represent conditions in Illinois or those for all Illinois programs. It is recommended that this protocol continue to be updated as further longer term and Illinois-specific evaluations are undertaken.

Illinois Statewide Technical Reference Manual for Energy Efficiency Version 5.0

Attachment A

Illinois Statewide Net-to-Gross Methodologies

**Effective for Evaluation:
June 1st, 2016**

All NTG data collection and analysis activities for the program types covered by this document that start after the effective date, June 1, 2016, shall conform to the NTG methods set forth herein.

Attachment A: Illinois Statewide Net-to-Gross Methodologies

1.1 Policy Context for this Information

The Illinois Evaluation Teams (ADM Associates, Cadmus Group, Itron, Navigant Consulting, Opinion Dynamics, Ridge & Associates) are working with the Illinois Stakeholder Advisory Group (SAG) to create an Illinois Statewide Net-to-Gross (NTG) Methodologies document (IL-NTG Methods). The IL-NTG Methods document is included as an attachment to the Illinois Statewide Technical Reference Manual for Energy Efficiency (IL-TRM). Through five different dockets, the Illinois Commerce Commission (ICC) has directed the Evaluation Teams to compile and formalize standard NTG methods for use in Illinois energy efficiency (EE) evaluation, measurement, and verification (EM&V) work. The ICC EE dockets are shown in the following table.

Table 1-1. ICC Energy Efficiency Dockets

ICC Order Docket No. and Date	Program Administrator	NTG Discussion – Order Pages	ICC Link
13-0495 (1/28/14)	Commonwealth Edison Company (ComEd)	129-130	http://www.icc.illinois.gov/downloads/public/docket/367591.pdf
13-0498 (1/28/14)	Ameren Illinois Company (Ameren)	167, 171	http://www.icc.illinois.gov/downloads/public/docket/367603.pdf
13-0499 (1/28/14)	Illinois Department of Commerce & Economic Opportunity (Department of Commerce)	20, 23, 49	http://www.icc.illinois.gov/downloads/public/docket/367581.pdf
13-0549 (5/20/14)	Nicor Gas Company (Nicor)	41-42, 78	http://www.icc.illinois.gov/downloads/public/docket/378494.pdf
13-0550 (5/20/14)	North Shore Gas Company (North Shore Gas) and The Peoples Gas Light and Coke Company (Peoples Gas) (collectively, PG&NSG)	54-55, 66	http://www.icc.illinois.gov/downloads/public/docket/378495.pdf

To provide clarity to the ICC directives, the relevant section on IL-NTG Methods is shown in its entirety from the Nicor Gas Order (Docket No. 13-0549). The Nicor Gas Order provides the most detail on the ICC NTG directive in comparison to the other EE orders. The Nicor language is as follows:

The Commission believes that Staff’s recommendations concerning Commission adoption of consistent statewide net-to-gross methodologies (“IL-NTG Methods”) for use by the evaluators are reasonable and will aid in future evaluation of the energy efficiency programs. To help ensure the independence of the evaluators, to improve efficiency in the evaluation process, and to ensure programs across the state as delivered by the various Program Administrators can be meaningfully and consistently evaluated, the Commission hereby adopts Staff’s recommendation that consistent IL-NTG Methods be established for use in the evaluations of comparable energy efficiency programs offered by different Illinois Program Administrators. The Commission notes that Section 8-104(k) of the Act encourages statewide coordination and consistency between the gas and electric energy efficiency programs and Staff’s proposal would help ensure consistency in the evaluation of program performance. The Commission notes that this directive is not to create entirely “new” NTG methodologies for every energy efficiency program, but rather to assess NTG methodologies and survey instruments that have been used to evaluate energy efficiency programs offered in Illinois, and to compile the most justifiable and well-

vetted methodologies (or potentially combine certain components from the existing approaches to better represent the most justifiable and well-vetted method consistent with best practices) in an attachment to the Updated IL-TRM that would get submitted to the Commission for approval. The Commission notes that the IL-NTG Methods will be flexible and adaptable to multiple program designs and budgets and tailored to appropriately assess the specifics of each of the Program Administrators' energy efficiency programs, consistent with standard NTG methodologies adopted in other states that were filed in this proceeding. The Commission agrees with Staff that in the interest of efficiency, the current program evaluators should take the lead in compiling and formalizing standard methodologies for NTG in Illinois taking into consideration SAG input. Because the existing Plan 1 evaluators are under contract with the Company for the evaluation of the program year three energy efficiency programs, it is appropriate for these existing evaluators to work on and complete the compilation of the IL-NTG Methods over the next year. The Commission recognizes that each year considerable time may be spent vetting NTG methodologies for each program evaluation separately for each utility under the existing evaluation plan review practices; adoption of IL-NTG Methods would save on these limited evaluation resources by having a common reference document for the evaluators to use in estimating net savings for Illinois.

The Commission hereby directs the Company to require its evaluators to collaborate with the other Illinois evaluators and the SAG to use best efforts to reach consensus on the approaches used in assessing NTG in particular markets for both residential and non-residential energy efficiency programs in a manner consistent with the direction described herein. (Pages 41-42)

- (16) Northern Illinois Gas Company shall require its evaluators to collaborate with the other Illinois evaluators and the SAG to reach consensus on the most defensible and well-vetted methodologies for assessing net-to-gross ratios in particular markets for both residential and non-residential energy efficiency programs in a manner consistent with the direction provided herein;
- (17) ICC Staff shall file the agreed-upon consensus statewide NTG methodologies with the Commission as an attachment to the Updated IL-TRM, and if consensus is not reached on a certain component of the statewide NTG methodologies, that particular non-consensus component should be submitted in a manner consistent with the approach used for non-consensus IL-TRM Updates; (Page 78)

1.2 Programs Currently Covered in this Document

This document is intended to cover the majority of residential and non-residential programs offered in Illinois.²⁸ Programs covered as of the writing of this document are listed in tables at the beginning of Section 3: Commercial, Industrial, and Public Sector Protocols and Section 4: Residential and Low Income Sector Protocols. If the design of a given program changes significantly, then it may mean that the NTG protocol listed for that program in this document is no longer appropriate. If that happens, the evaluator should follow the procedures outlined below under Section 1.4: Diverging from the IL-NTG Methods.

This document will be updated over time to incorporate new programs and to reflect recommended changes to existing methodologies. All NTG data collection and analysis activities for the program types covered by this document that start after the effective date, June 1, 2016, shall conform to the NTG methods set forth herein.

²⁸ Evaluation reports on those programs can be found at <http://www.ilsag.info/evaluation-documents.html>.

1.3 Updating the IL-NTG Methods

This attachment is part of the IL-TRM and follows the timeline for updating of the IL-TRM, as specified in the IL-TRM Policy Document.²⁹ In general, the following will take place:

- Updates will occur annually.
- Any changes to the IL-NTG Methods document will be circulated to the full SAG, and SAG participants will have a ten business day review process.
- Updates will be discussed within the SAG and completed by March 1.
- The ICC Staff will then submit a Staff Report (with the consensus Updated IL-TRM attached) to the Commission with a request for expedited review and approval.

1.4 Diverging from the IL-NTG Methods

The NTG methods for the programs outlined in this document are partially binding. The criteria for deviating from the IL-NTG Methods document are set forth below. In all cases, the evaluators (or any interested stakeholder) submits the proposed deviation to the full SAG for a ten business day SAG review and comment period. In the event of an objection by a SAG participant, efforts may be made to see if consensus can be reached on the proposed deviation in a subsequent monthly SAG meeting. In this case, a final opportunity for SAG review and comment to the proposed deviation will be provided following the SAG meeting.

Evaluators may modify the approaches described in this document if the following three conditions have been satisfied:

1. Evaluators must explicate within the annual evaluation research plan (or another document) how specific items in the proposed modified NTG method will diverge from what is written in this document. Evaluators must justify why the divergence is appropriate.
2. Prior to the use of the modified NTG method for a particular program, evaluation teams must be in agreement on the use and execution of the modified NTG method.
3. Any objection from SAG participants regarding the proposed modified NTG method is resolved.

Evaluators may test alternative methods of estimating NTG for a particular program in addition to the NTG methods outlined in this document, if the following three conditions have been satisfied:

1. Evaluators must explicate within the annual evaluation research plan (or other document) the proposed alternative NTG method. Evaluators must explain why the proposed alternative NTG method might be superior to the NTG methods outlined in this document for the particular program. Evaluators must discuss the foundation for expecting that the proposed alternative NTG method is likely to produce meaningful results.
2. Prior to the use of the alternative NTG method for a particular program, evaluation teams must be in agreement on the key details of the approach for implementing the alternative NTG method.
3. Any objection from SAG participants regarding the proposed alternative NTG method gets resolved.

When performing alternative NTG methods for a particular program, the choice of methods may vary across the state. For example, if ComEd's evaluator chooses to test Methods 1 and 2 for a particular program, Ameren's and

²⁹ Policy Document for the Illinois Statewide Technical Reference Manual for Energy Efficiency. October 25, 2012. <http://www.icc.illinois.gov/downloads/public/IL%20TRM%20Policy%20Document.pdf>

Department of Commerce's evaluators do not also have to perform Methods 1 and 2 for a similar program.

Several sections of this attachment provide example questions that can be used to collect the data required in the NTG algorithms. Adjustments to refine specific question wording, e.g., to better reflect the design of the evaluated program, do not constitute divergence from the IL-NTG Methods.

1.5 Procedure for Non-Consensus Items

Non-consensus items that arise during the development and updating of the IL-NTG Methods document will be handled in substantially the same way as non-consensus IL-TRM Updates are addressed. The approach to be used is as follows.

- Once the Illinois NTG Working Group³⁰ has progressed as far as they can on the methodology, and it has been found that there is non-consensus on a specific Net-to-Gross Methods topic or procedure, the Illinois NTG Working Group shall submit to the ICC Staff and the SAG's Technical Advisory Committee (TAC) a Comparison Exhibit of Non-Consensus Net-to-Gross Methods topics/procedures within two weeks after the Illinois NTG Working Group has failed to reach consensus. The TAC will then deliberate on the issue with a goal of reaching consensus.
- If consensus does not emerge in the TAC regarding a particular Net-to-Gross Methods topic or procedure, the Comparison Exhibit of Non-Consensus NTG Methods topics/procedures is then sent to the full SAG for their deliberations and input. The SAG provides a forum where experts on all sides of the contested issue can present their expert opinions in an effort to inform parties of the contested issue and to also facilitate consensus.
- If the full SAG is unable to reach consensus, the non-consensus item will be referred to the ICC for resolution at the time of the IL-TRM Update proceeding. After receipt of the Comparison Exhibit of Non-Consensus Net-to-Gross Methods topics/procedures, the ICC Staff will submit a Staff Report to the Commission to initiate a proceeding separate from the consensus IL-TRM Update proceeding to resolve the non-consensus Net-to-Gross Methods topics/procedures.

³⁰ The Illinois NTG Working Group consists primarily of the subset of Evaluators deliberating on NTG methodologies; however, any interested party may participate in the Illinois NTG Working Group.

2 Attribution in Energy Efficiency Programs in General

One of the most difficult aspects of evaluation, and not just within evaluation of energy efficiency programs, is attributing results to a program. Attribution provides credible evidence that there is a causal link between the program activities and the outcomes achieved by the program. Attribution research estimates the difference between the outcomes and those that would have occurred absent the program (i.e., the counterfactual). Put in research terms, evaluators must reject the null hypothesis of no causality through probabilistic statements (e.g., “strong evidence”; “high probability”). As such, it is important to realize that the concept of the counterfactual cannot be proven with certainty. So even though the NTG ratio is a single value, conceptually it is a probabilistic statement.³¹ One of the main academics within evaluation stated that there is a “...total and inevitable absence of certain knowledge [arising] from the methods social scientists use” when assessing the counterfactual. (Shadish, et al., 2002) This statement is not about poor methods, but about the counterfactual itself. Because programs work with people and are usually not a laboratory experiment that can be replicated over and over³² to find out what actions people would have taken absent an intervention, one would need a time machine to take people back in time and not provide the program. Since time machines do not exist, evaluators have developed methods that approximate the counterfactual to the best of their ability.

2.1 Definitions

For energy efficiency programs, evaluators differentiate between savings at a “gross” and “net” level as described below in the short set of relevant definitions. These definitions are not all encompassing or meant to restrict evaluation in any way, but to provide context before additional detail is provided in later sections. Research to determine attribution occurs to allow for a better understanding of the net level of savings.

Table 2-1. Definitions

Concept	Term	Definition
Consumers	Nonparticipant	Any consumer who was eligible but did not participate in the subject efficiency program, in a given program year.
	Participant	A consumer who received a service offered through the subject efficiency program, in a given program year; also called program participant. The term “service” is used in this definition to suggest that the service can be a wide variety of inducements, including financial rebates, technical assistance, product installations, training, energy efficiency information, or other services, items, or conditions. Each evaluation plan should define “participant” as it applies to the specific evaluation.
Gross Impacts	Gross Impacts	The change in energy consumption and/or demand that results directly from program-related actions taken by participants in an energy efficiency program, regardless of why they participated.

³¹ A probabilistic statement is not the same as the confidence and precision information calculated based on sampling theory.

³² However, a small number of program designs do lend themselves to experimental or quasi-experimental designs that allow for regression analysis of net impacts.

Concept	Term	Definition
Attribution of Impacts	Net Impacts	The change in energy consumption and/or demand that is attributable to a particular energy efficiency program. This change in energy use and/or demand may include, implicitly or explicitly, consideration of factors such as free ridership, participant and nonparticipant spillover, and induced market effects. These factors may be considered in how a baseline is defined (e.g., common practice) and/or in adjustments to gross savings values.
	Net-to-Gross Ratio	A factor representing net program savings divided by gross program savings that is applied to gross program impacts to convert them into net program impacts. The factor itself may be made up of a variety of factors that create differences between gross and net savings, commonly including free riders and spillover. The factor can be estimated and applied separately to either energy or demand savings. Note that the net-to-gross ratio (NTGR) = $((1 - \text{Free Ridership}) + \text{Participant Spillover} + \text{Nonparticipant Spillover})$.
	Core NTGR	1-Free Ridership
	Free Rider	A program participant who would have implemented the program's measures or practices in the absence of the program. Free riders can be: (1) total, in which the participant's activity would have completely replicated the program measure; (2) partial, in which the participant's activity would have partially replicated the program measure; or (3) deferred, in which the participant's activity would have partially or completely replicated the program measure, but at a future time.
	Spillover	<p>Reductions in energy consumption and/or demand caused by the presence of an energy efficiency program, beyond the program-claimed gross savings of the participants. There can be participant and/or nonparticipant spillover.</p> <p><i>Participant spillover</i> (PSO) is the additional energy savings that occur as a result of the program's influence when a program participant independently installs incremental energy efficiency measures or applies energy-saving practices after having participated in the energy efficiency program. There are several general categories of participant spillover:</p> <ul style="list-style-type: none"> • <i>Inside spillover</i> (ISO): Occurs when program participants implement additional program-induced energy efficiency measures at the program project site. • <i>Outside spillover</i> (OSO): Occurs when program participants implement program-induced efficiency measures at other sites within the Program Administrator's service territory at which program project measures were not implemented. • <i>Like spillover</i>: Occurs when program participants implement program-induced efficiency measures of the same type as those implemented through the program. Like spillover can occur at the program project sites (ISO) or at other sites within the Program Administrator's service territory (OSO). • <i>Unlike spillover</i>: Occurs when program participants implement

Concept	Term	Definition
		<p>program-induced efficiency measures of a different type from those implemented through the program. Unlike spillover can occur at the program project sites (ISO) or at other sites within the Program Administrator's service territory (OSO).</p> <p><i>Nonparticipant spillover (NPSO)</i> refers to energy savings that occur when a program nonparticipant installs energy efficiency measures or applies energy savings practices as a result of a program's influence.</p>
Markets	Market	The commercial activity (e.g., manufacturing, distributing, buying, and selling) associated with products and services that affect energy use.
	Market Effects	A change in the structure of a market or the behavior of participants in a market that is reflective of an increase (or decrease) in the adoption of energy efficient products, services, or practices and is causally related to market interventions (e.g., programs). Examples of market effects include increased levels of awareness of energy-efficient technologies among customers and suppliers, increased availability of energy-efficient technologies through retail channels, reduced prices for energy-efficient models, build-out of energy-efficient model lines, and—the end goal— increased market shares for energy-efficient goods, services, and design practices.
	Market Assessment	An analysis that provides an assessment of how and how well a specific market or market segment is functioning with respect to the definition of well-functioning markets or with respect to other specific policy objectives. A market assessment generally includes a characterization or description of the specific market or market segments, including a description of the types and number of buyers and sellers in the market, the key actors that influence the market, the type and number of transactions that occur on an annual basis, and the extent to which market participants consider energy efficiency an important part of these transactions. This analysis may also include an assessment of whether a market has been sufficiently transformed to justify a reduction or elimination of specific program interventions (or whether continued or even increased intervention is necessary). Market assessment can be blended with strategic planning analysis to produce recommended program designs or budgets. One particular kind of market assessment effort is a baseline study, or the characterization of a market before the commencement of a specific intervention in the market for the purpose of guiding the intervention and/or assessing its effectiveness later.

Sources: State and Local Energy Efficiency Action Network. 2012. Energy Efficiency Program Impact Evaluation Guide. Prepared by Steven R. Schiller, Schiller Consulting, Inc., www.seeaction.energy.gov; Violette and Rathbun 2014. The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures, Chapter 23: Estimating Net Savings: Common Practices, <http://www.nrel.gov/docs/fy14osti/62678.pdf>.

2.2 Spillover-Specific Issues

Some issues related to spillover are applicable for both residential and non-residential programs and are discussed in this section.

2.2.1 Measure Costs

In order to facilitate analysis of program Total Resource Cost (TRC), estimates of the total incremental measure cost (IMC) at the program level must be developed. IMC values are available for most IL-TRM measures and can be summed to the program level. However, the IMC values for spillover measures could also be estimated and added to this total. The problem is that IMC values for spillover measures can be difficult to estimate. When the magnitude of the savings justifies the effort to estimate the total IMC for spillover measures, the following approaches should be used.

- In cases where the evaluator believes the spillover measure incremental costs are not materially different from the rebated measure incremental costs, the evaluator may multiply the IMC for the rebated measure by the spillover rate to derive the IMC for the spillover measure.
- In cases where the evaluator believes the spillover measure incremental costs are materially different from the installed measure incremental costs (e.g., installation of measures that have no efficiency levels), the evaluator should use the estimated incremental project costs as the IMC for the spillover measure.

Normally, the sample-based estimates of IMCs for spillover measures should be extrapolated to the program level using sample weights. Then the total IMCs for rebated measures and the total IMCs for spillover measures should be summed and used in the TRC calculation.

For measures characterized by the IL-TRM, measure effective useful life (EUL) estimates should be based on the IL-TRM. For measures not characterized by the IL-TRM, evaluator can use either the EUL for similar measures or best professional judgment. In either case, the evaluator must provide the rationale for their choices.

3 Commercial, Industrial, and Public Sector Protocols

The table below lists Illinois non-residential programs and the free ridership protocol applicable to each program.³³ If the design of a given program changes significantly, then it may mean that the NTG protocol listed for that program in this document is no longer appropriate. If that happens, the evaluator should follow the procedures outlined in Section 1.4: Diverging from the IL-NTG Methods. Note that the Core Non-Residential Spillover protocol described in Section 3.2 is generally applicable to most of these programs.

Table 3-1. Commercial, Industrial, and Public Sector Programs

Program Administrator	Free Ridership Protocol	Program Name
Ameren Illinois	3.1 Core Non-Residential Protocol	C&I Custom
		C&I Standard
	3.3 Small Business Protocols	Small Business Direct Install
	3.5 Study-Based Protocol	C&I Retro-Commissioning
ComEd	3.1 Core Non-Residential Protocol	BILD / Midstream
		Custom Incentive
		Savings through Efficient Products (STEP)
		Standard Incentive
	3.3 Small Business Protocol	Agricultural EE Program (CoAg)
		CLEAResult School DI
		DCV – Matrix Demand- Based Fan Control
		EE Technologies to Address Peak Load in Assisted Living and Senior Housing
		Luminaire Level Lighting Control
		Matrix K through 12 Private Schools DI
		Rural Small Business EE Kits
		Small Business Energy Services
		Small Commercial Lit Signage
		Small Commercial HVAC Tuneup
	3.4 C&I New Construction Protocol	C&I New Construction
		New Construction – Small Buildings
	3.5 Study-Based Protocol	Data Centers
		Enhanced Building Optimization Program

³³ The “Free Ridership Protocol Name” in the second column of the table refers to the numbered sections in this document, e.g., “3.3 Small Business Protocol.”

Program Administrator	Free Ridership Protocol	Program Name
		Industrial Systems Optimization
		Retrocommissioning
		Strategic Energy Management
	3.5 Study-Based Protocol or 5.1 Behavioral Protocol	Power TakeOff – Small Business MBCx
	5.1 Behavioral Protocol	Agentis C&I Behavioral Program
		Root3
	NTG = 1	LED Streetlighting
Department of Commerce	3.1 Core Non-Residential Protocol	Public Sector Custom
		Public Sector Custom - CHP Component
		Public Sector Natural Gas Boiler Systems Efficiency
		Public Sector Standard
		Savings through Efficient Products
	3.4 C&I New Construction Protocol	Public Sector New Construction
	3.5 Study-Based Protocol	Public Sector Retro-Commissioning
	3.6 Training and Technical Assistance Protocol	Building Operator Certification
		Energy Assessment and New Construction Design Assistance
		Performance Contracting
	5.2 Code Compliance Protocol	Building Energy Code Compliance
Nicor	N/A	Emerging Technology
	3.1 Core Non-Residential Protocol	Business Custom (Except Retro-Commissioning)
		Business Energy Efficiency Rebate
		Combined Heat and Power (CHP)
	3.3 Small Business Protocol	Small Business Energy Savings
	3.4 C&I New Construction Protocol	Business New Construction
	3.5 Study-Based Protocol	Business Custom (Retro-Commissioning component)
		Strategic Energy Management (SEM)
Peoples Gas/ North Shore Gas	3.1 Core Non-Residential Protocol	C&I Custom
		C&I Direct Install
		C&I Prescriptive
	3.3 Small Business Protocol	SB Custom
		SB Direct Install & Assessment

Program Administrator	Free Ridership Protocol	Program Name
		SB Partner Trade Ally
		SB Prescriptive
	3.4 C&I New Construction Protocol	C&I New Construction (Joint)
	3.5 Study-Based Protocol	C&I Gas Optimization
		MF Gas Optimization
		Retro-Commissioning (Joint)
All	5.2 Code Compliance Protocol	Statewide Codes Collaborative

3.1 Core Non-Residential Protocol

3.1.1 Core Non-Residential Free Ridership Protocol

Key considerations and guidelines for estimation of free ridership under this Core Non-Residential Free Ridership (FR) protocol are listed below:

- Multiple Questions:** Evaluators will use program participant responses to multiple survey questions as inputs to the free ridership calculation algorithm. Evaluators will not use the response to a single question to establish a survey respondent as either a complete free rider or a complete non-free rider.
- Program and Non-Program Factors:** Evaluators will administer survey questions to obtain respondent ratings on a numeric scale of the impact, influence, or importance on the decision to implement energy efficiency measures or take energy efficiency actions. A series of questions will focus on factors that the evaluator determines are a function of the program. Such program factors may, for instance, include availability of the program incentive, technical assistance from program staff, program staff recommendations, Program Administrator marketing materials, and endorsement or recommendation by Program Administrator account manager or program partner staff. Evaluators will also administer a series of questions to obtain respondent ratings, on a numeric scale of the impact, influence, or importance on the decision to implement energy efficiency measures, of different factors that the evaluator determines are not a function of the program. Such non-program factors may include, for example, previous experience with the measure, standard business or industry practice, and organizational policy or guidelines.
- Vendor Recommendations:** Vendor recommendations may also be a program factor to the extent that such recommendations are a function of the program. Vendors include trade allies, contractors, distributors, suppliers, and other market actors involved in the selection and installation of program-incented equipment on behalf of the participant. The evaluator may administer survey questions to vendors to verify their involvement with participant projects and to obtain their ratings—on a numeric scale—of the impact, influence, or importance of the program on the decision to recommend the energy efficiency measures to the program participant.
- Consistency Checks:** Evaluators should administer survey questions as checks on the consistency of responses associated with a core free ridership assessment methodology. Evaluators may also reference available quantitative and qualitative data, including consistency check data, to perform documented modifications to individual free ridership estimates resulting from the application of a core free ridership assessment methodology.

3.1.1.1 Core Free Ridership Scoring Algorithm

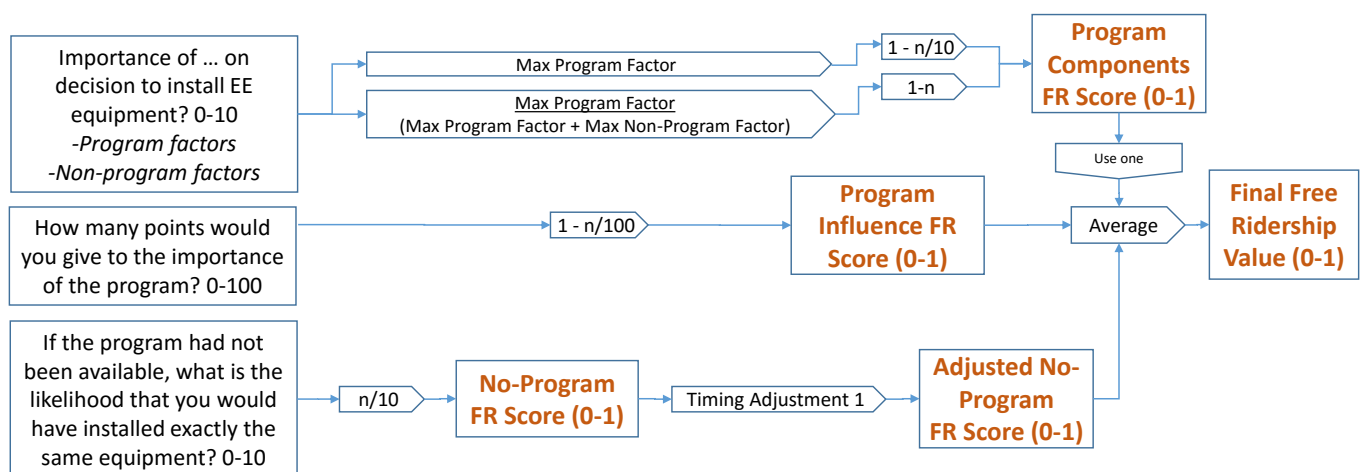
The Core Non-Residential FR protocol combines three scores that test different ways of approaching free ridership: the Program Components FR Score, the Program Influence FR Score, and the No-Program FR Score. The three scores are combined to calculate the final free ridership value. Three options for combining the three scores are shown graphically in Figure 3-1, Figure 3-2, and Figure 3-3.

This protocol designates an algorithm that includes inputs with alternative specifications. Specifically, as described below, the Program Components FR Score has two alternative specifications, and there are three alternative specifications to account for the impact of the program on project timing (referred to as “deferred free ridership”). Further, the way in which the algorithm inputs are combined and weighted varies, based on the selected specification. Evaluators will calculate free ridership using each possible combination of designated input specifications—resulting in six estimates of free ridership—and will select one of these combinations of input specifications for purposes of calculating the annual incremental energy savings for comparing to the legislated goal.³⁴

Evaluators will submit participant survey and net savings analysis data to the Illinois NTG Working Group. The group will analyze these data for the purpose of further refining the protocol and potentially reducing the number of alternative algorithm input specifications.

Figure 3-1. Core Free Ridership Algorithm 1

(Program Components FR Score + Program Influence FR Score + (No-Program FR Score * Timing Adjustment 1)) / 3



³⁴ As defined in 220 ILCS 5/8-103 and 220 ILCS 5/8-104.

Figure 3-2. Core Free Ridership Algorithm 2

$((\text{Program Components FR Score} + \text{Program Influence FR Score} + \text{No-Program FR Score}) / 3) * \text{Timing Adjustment 2}$

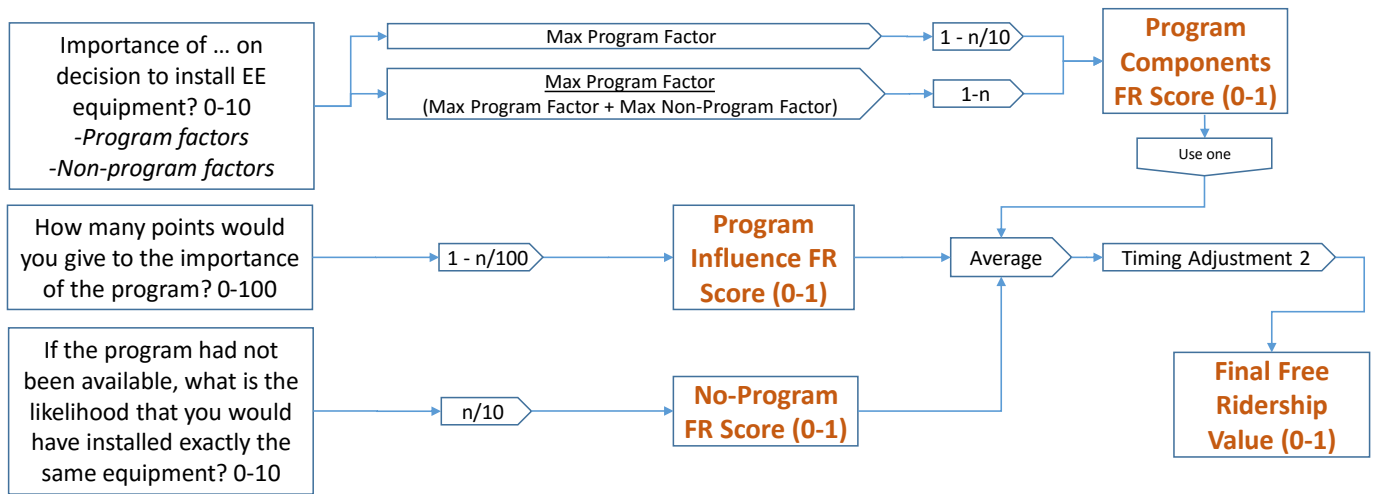
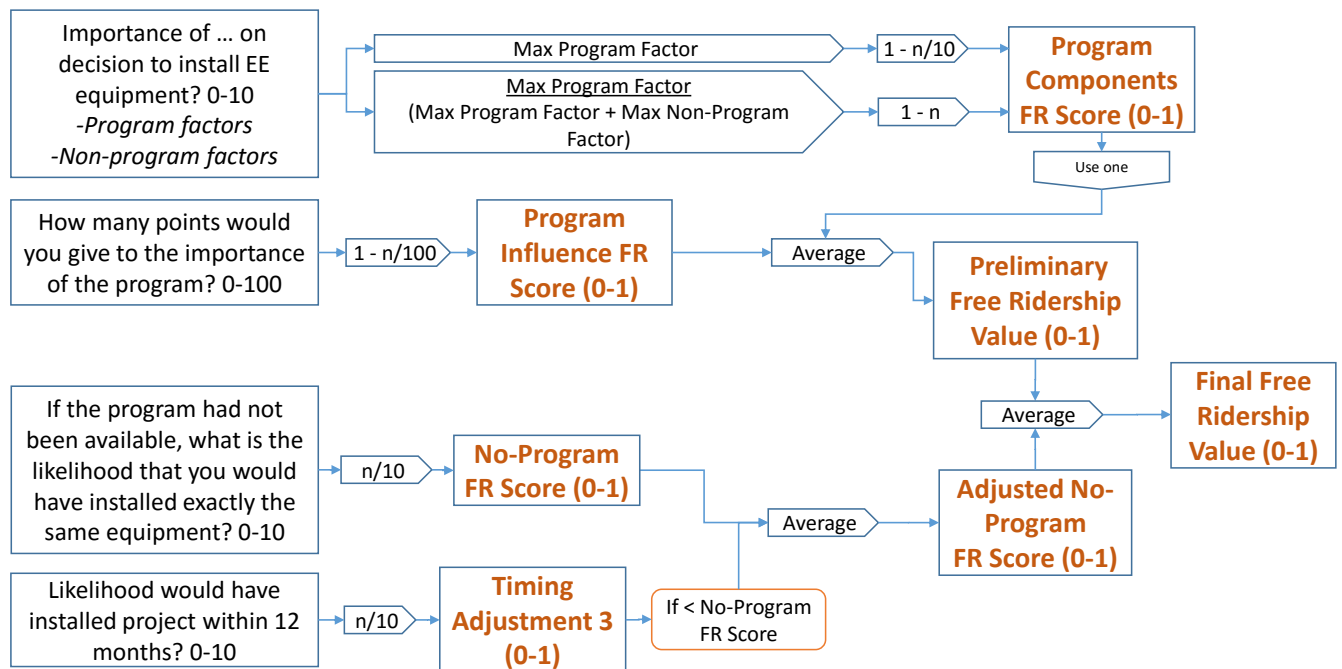


Figure 3-3. Core Free Ridership Algorithm 3

$((\text{Program Components FR Score} + \text{Program Influence FR Score}) / 2 + (\text{No-Program FR Score} + \text{Timing Adjustment 3}) / 2) / 2$



3.1.1.1.1 Program Components FR Score

Evaluators will administer survey questions to obtain participants' rating of the importance of various factors on the decision to implement energy efficiency measures. The numeric scales shall range from 0 to 10, where 0 means "not at all important" and 10 means "extremely important". The various factors referenced in the survey will include those that the evaluator determines are program factors and non-program factors that could potentially impact the participant decision making process. A participant rating shall be obtained for each relevant program and non-program factor.

Evaluators will calculate the “Program Components FR Score” for each survey respondent in two ways (shown in the top center of the three flow charts). These scores can range from 0 (no free ridership) to 1 (full free rider).

- 1) Equal to $1 - ([\text{Maximum Program Factor Rating}]/10)$.
- 2) Equal to $1 - ([\text{Maximum Program Factor Rating}]/([\text{Maximum Program Factor Rating}] + [\text{Maximum Non-Program Factor Rating}]))$.

Evaluators may use one of the two above-referenced calculation methods to develop a Program Components FR Score to use as an input to the algorithm for calculation of program-level free ridership. Evaluators must document in EM&V reporting why the selected option was chosen. Evaluators should also test the other approach and report on the effect of their choice on the results.

Evaluation reports should list all factors considered program and non-program factors. Evaluators must document why factors were treated as program factors or non-program factors.

3.1.1.1.2 Program Influence FR Score

Evaluators will administer a survey question that asks respondents to quantify the importance of the program on the decision to implement energy efficiency measures relative to the importance or impact of non-program factors. Respondents will be asked to allocate a total of 100 points to the program and to non-program factors. The points allocated to the program by the participants are the “Program Points.” Evaluators will calculate the “Program Influence FR Score” as $1 - (\text{Program Points}/100)$. This score can range from 0 (no free ridership) to 1 (full free rider).

3.1.1.1.3 No-Program FR Score

Evaluators will administer a counterfactual likelihood survey question to obtain respondent ratings on a 0 to 10-point numeric scale (where 0 means “not at all likely” and 10 means “extremely likely”) of the likelihood of the respondent to implement the exact same energy efficiency measures in the absence of the program. Evaluators will calculate the “No-Program FR Score” as the numeric score of the likelihood of the respondent to implement specified energy efficiency measures in the absence of the program divided by 10. This score can range from 0 (no free ridership) to 1 (full free rider).

Note that under two of the three deferred free ridership specifications (see next subsection), a timing adjustment is applied to the “No-Program FR Score.” Under these two specifications, the resulting score is referred to as the “Adjusted No-Program FR Score.”

3.1.1.1.4 Timing and Deferred Free Ridership

Evaluators will ask about the likely timing of measure installation in the absence of the program in two different ways. This is referred to as the counterfactual timing questions since the evaluators are asking the respondent to speculate on what might have happened within particular timeframes.

The first question will present a series of date ranges (e.g., within one year, between 12 months and 2 years, etc.) and ask the respondent to pick one representing their best estimate of when the measure would have been implemented in the absence of the program. The free ridership algorithm uses the midpoint of each date range, referred to as “Number of Months Expedited” below. For respondents that report accelerated adoption due to the program, this variable can take on values from 6 to 48 months.

The second question will prompt the respondent to use a 0 to 10-point numeric scale to report the likelihood, in the absence of the program, of implementing the same measure within 12 months of when it was actually implemented. This is the “Likelihood of Implementing within One Year” in the formulas below.

Evaluators will use the Likelihood of Implementing within One Year and/or the Number of Months Expedited variables to calculate three alternative ways of accounting for deferred free ridership:

- 1) Calculate *Timing Adjustment 1* as equal to:

$$1 - (\text{Number of Months Expedited} - 6)/42$$

Timing Adjustment 1 is multiplied by the No-Program FR Score; it can range from 0 (full deferred free ridership) to 1 (no deferred free ridership). The application of Timing Adjustment 1 is shown in Figure 3-1.

- 2) Calculate *Timing Adjustment 2* as equal to:

$$1 - ((\text{Number of Months Expedited} - 6)/42) * ((10 - \text{Likelihood of Implementing within One Year})/10)$$

Timing Adjustment 2 is multiplied by the average of the Program Components FR Score, the Program Influence FR Score, and the No-Program FR Score; it can range from 0 (full deferred free ridership) to 1 (no deferred free ridership). The application of Timing Adjustment 2 is shown in Figure 3-2.

- 3) Calculate *Timing Adjustment 3* as equal to:

$$\text{Likelihood of Implementing within One Year}/10$$

Timing Adjustment 3 is averaged with the No-Program FR Score; it can range from 0 (full deferred free ridership) to 1 (no deferred free ridership). The application of Timing Adjustment 3 is shown in Figure 3-3.

How these timing adjustments are accounted for in the calculation of the Final FR Value is described below in the subsection “3.1.1.2 Construction of Core Free Ridership Value.”

3.1.1.1.5 Consistency Checks

Respondents may be asked one or more questions to facilitate understanding and potentially reconciling apparently inconsistent responses. Some questions may be asked of all respondents; others may be asked when previous answers appear inconsistent. Evaluators should report on the amount of inconsistency encountered and on the resolution to inform future protocol revisions. Three consistency checks are outlined below.

Program Influence/Program Components Consistency Check

A Program Influence/Program Components consistency check is triggered when the following conditions are met:

- 1) The number of Program Points (supporting calculation of the Program Influence FR Score) is greater than 70; and
- 2) no program factor is rated greater than 2.

A Program Influence/Program Components consistency check is also triggered by the following conditions being met:

- 1) The number of Program Points (supporting calculation of the Program Influence FR Score) is less than 30; and
- 2) at least one program factor is rated greater than 7. In this instance, the highest-rated program factor(s) with a rating of greater than 7 will be referenced in the consistency check question.

Program Components/No-Program Consistency Check

A Program Components/No-Program consistency check is triggered when the following conditions are met:

- 1) The likelihood of installing the exact same equipment without the program (supporting calculation of the No-Program FR Score) is greater than 7; and
- 2) at least one program factor is rated greater than 7.

A Program Components/No-Program consistency check is also triggered when the following conditions are met:

- 1) The likelihood of installing the exact same equipment without the program (supporting calculation of the No-Program FR Score) is less than 3; and
- 2) no program factor is rated greater than 2.

Timing of Installation Decision/Level of Program Attribution Consistency Check

The survey should contain a question to ask whether the respondent learned about the program after finalizing project specifications, including, where applicable, equipment efficiency level and number of units. The Timing of Installation Decision/Level of Program Attribution consistency check is triggered by the following conditions being met:

- 1) A respondent learned about the program after finalizing project specifications; and
- 2) any of the following occur:
 - a) the number of Program Points (supporting calculation of the Program Influence FR Score) is greater than 70;
 - b) the likelihood of installing the exact same equipment without the program (supporting calculation of the No-Program FR Score) is less than 3; or
 - c) at least one program factor is rated greater than 7.

When the Timing of Installation Decision/Level of Program Attribution consistency check is administered, if the respondent rating of the importance of the vendor on the decision to implement the project is greater than 7, then an open-ended question will be triggered to obtain information regarding the role the vendor played in the participant decision to implement the project.

3.1.1.2 Construction of Core Free Ridership Value

This protocol designates an algorithm that includes inputs with alternative specifications. Specifically, as described above, the Program Components FR Score input has two alternative specifications and the Deferred Free Ridership input has three alternative specifications. Evaluators will calculate free ridership using each possible combination of designated input specifications – resulting in six estimates of free ridership – and will select one of these combinations of input specifications for purposes of calculating the annual incremental energy savings for comparing to the legislated goal. Evaluators will present the results of all six estimates of free ridership in EM&V reporting.

Evaluators will calculate free ridership values in the following three ways in order to account for the variable ways of accounting for deferred free ridership outlined above:

- 1) Core FR Algorithm 1 = $AVERAGE([Program\ Components\ FR\ Score], [Program\ Influence\ FR\ Score], [No-Program\ FR\ Score * Timing\ Adjustment\ 1])$
- 2) Core FR Algorithm 2 = $AVERAGE([Program\ Components\ FR\ Score], [Program\ Influence\ FR\ Score], [No-Program\ FR\ Score]) * Timing\ Adjustment\ 2$
- 3) Core FR Algorithm 3 = $AVERAGE([(Program\ Components\ FR\ Score + Program\ Influence\ FR\ Score)/2], [MINIMUM((No-Program\ FR\ Score + Timing\ Adjustment\ 3)/2, No-Program\ FR\ Score)])$

For each of the three ways, the Program Components FR Score is calculated in two ways:

- a) $1 - ([Maximum\ Program\ Factor\ Rating]/10)$
- b) $1 - ([Maximum\ Program\ Factor\ Rating]/([Maximum\ Program\ Factor\ Rating] + [Maximum\ Non-Program\ Factor\ Rating]))$

The three Core FR Algorithms listed above are graphically presented in Figure 3-1, Figure 3-2, and Figure 3-3, respectively.

3.1.1.3 Vendor Influence in the Free Ridership Calculation

3.1.1.3.1 Treatment of Participant's Rating of Vendor in the Program Components FR Score of the Core FR Algorithm

The Program Components FR Score of the participant Core FR algorithm is based on participant ratings of program and non-program factors. Vendors³⁵ often receive a high rating for their influence on the participant's decision to install the efficient measure. To implement the Core FR algorithm, the evaluator needs to decide whether the vendor rating should be considered a program factor or a non-program factor. This section outlines three scenarios for the treatment of the participant's rating of a vendor in the Program Components FR Score of the Core FR algorithm.

Scenario #1: Vendors are automatically considered a program factor

The vendor is considered a program factor in the calculation of the Program Components FR Score in the FR algorithm if the program meets specific criteria, which could include the following:

1. Trade allies are an integral component of program delivery, as supported by program logic
2. The trade ally network consists of a limited number of Program Administrator-selected, pre-approved trade allies
3. Only trade allies can implement projects and submit applications on behalf of the customer
4. Trade allies complete signed agreements with the Program Administrator
5. Trade allies complete program-sponsored training

In these cases, the vendor is automatically considered a program factor, and no additional input from the vendor is needed regarding the customer's decision-making process related to the project. The participant's influence rating for the vendor goes directly into the Program Components FR Score algorithm as a program factor (if it is the highest rating given to any program factor).

Scenario #2: Vendors are considered a program factor if the program influenced their recommendation to implement the efficient project

For programs that have a trade ally network, but do not meet the conditions under Scenario #1 above, follow-up interviews with vendors may be used to determine if the vendor should be considered a program factor. To qualify for Scenario #2, a program's trade ally network should meet the following conditions:

1. Trade allies are registered with the program
2. Trade allies typically complete signed agreements with the Program Administrator
3. Trade allies complete program-sponsored training
4. Trade allies drive program participation, as supported by program logic

In these cases, if the size of the project warrants a greater level of effort, a follow-up interview with the vendor may be used to determine if the participant's rating of the vendor's influence should be included as a program factor. A follow-up interview is triggered under the following conditions:

1. The participant rated the influence of the vendor as 8, 9, or 10 (on a scale from 0 to 10)
2. The rating the participant gave to vendor influence is higher than any of the program factor ratings

If completed, the interview should include the following questions:

FR1a On a scale of 0 to 10 where 0 is NOT AT ALL IMPORTANT and 10 is EXTREMELY IMPORTANT, how important was the <PROGRAM>, including incentives as well as program services and information, in influencing your decision to recommend that <CUSTOMER> install the energy efficient <MEASURE> at this time?

³⁵ Vendors include trade allies, contractors, distributors, suppliers, and other market actors involved in the selection and installation of program-incented equipment on behalf of the participant.

- FR1b On the same scale, how important was your firm's past participation in an incentive or study-based program sponsored by <PROGRAM ADMINISTRATOR>?
- FR2 And using a 0 to 10 likelihood scale where 0 is NOT AT ALL LIKELY and 10 is EXTREMELY LIKELY, if the <PROGRAM>, including incentives as well as program services and information, had not been available, what is the likelihood that you would have recommended this specific <MEASURE> to <CUSTOMER>?
- FR3a Approximately, in what percent of projects did you recommend <MEASURE> BEFORE you learned about the <PROGRAM>?
- FR3b And approximately, in what percent of projects do you recommend <MEASURE> now that you have worked with the <PROGRAM>?

The interview will also include consistency checks, if the vendor provides inconsistent responses to these questions.

The vendor is viewed as a program factor and the rating the participant provided for the vendor goes into the Program Components FR Score algorithm as a program factor if, after consideration of any consistency checks:

1. The response to Q. FR1a or FR1b is 8, 9, or 10

OR

2. The response to Q. FR2 is 0, 1, or 2

OR

3. The difference between the responses to FR3b and FR3a is 80% or greater

If none of these conditions are met, the rating the participant provided for the vendor goes into the Program Components FR Score algorithm as a non-program factor.

In the event that an interview is not completed (e.g., the size of the project did not warrant a vendor interview or the vendor could not be reached), the evaluation reports should explain how the rating the participant provided for the vendor was treated. Guidelines for these situations may be added to this document in the future.

Scenario #3: Vendors are considered a non-program factor

For programs that do NOT have a trade ally network that meets the conditions under Scenario #2, vendors are considered a non-program factor. In these cases, the participant's rating of the vendor goes directly into the Program Components FR Score algorithm as a non-program factor (if it is the highest rating given to any non-program factor).

3.1.1.4 Public Sector Planning

The ICC order in Docket No. 11-0593 stated:

All parties, including DCEO, are cautioned that, with respect to a determination regarding "free ridership," the person or entity in question should have actual energy efficiency plans before they are to be considered to be "free riders," as opposed to persons who have some goal to be met in the distant future regarding energy efficiency products and services.³⁶

The NTG Working Group did not achieve consensus on an approach to address this issue and will continue to work toward consensus for a future version of this document. In the interim, evaluators may make documented adjustments to the free ridership analysis approach outlined in this section for the purpose of generating free ridership estimates that account for this order. They should present NTG results both with and without this adjustment.

³⁶ <https://www.icc.illinois.gov/downloads/public/edocket/371251.pdf>

3.2 Core Non-Residential Spillover Protocol

Spillover refers to energy savings associated with energy-efficient equipment installed by consumers who were influenced by an energy efficiency program, but without direct intervention (e.g., financial or technical assistance) from the program. Energy savings associated with spillover are not included in the Program Administrator's energy savings claim, nor are they formally tracked in the Program Administrator's databases.

To place the spillover protocols in context, we begin by defining the NTGR as:

$$\text{NTGR} = (1 - \text{Free Ridership Value} + \text{PSO Rate} + \text{NPSO Rate})$$

Where:

PSO Rate = Participant spillover rate

NPSO Rate = Nonparticipant spillover rate

The term (1-Free Ridership) is referred to as the Core NTGR for an efficiency program.

3.2.1.1 Core Participant Spillover Protocol

The Core Participant Spillover protocol is generally applicable to most commercial, industrial, and public sector programs.

3.2.1.1.1 Research Methods

Data collection approach. An initial determination of participant spillover may be made based on self-reported findings from surveys of program participants. At a minimum, surveys collecting data pertaining to participant spillover will obtain general information on the specific measures installed and information substantiating their attribution to an energy efficiency program. Research on the specific characteristics of the energy efficient equipment installed and the baseline and operating conditions needed to estimate savings may be done in one of two ways: 1) a detailed battery of measure specific questions may be administered as part of the initial survey; or 2) a separate in-depth follow-up interview may be conducted by the engineer or analyst responsible for the energy savings calculation. In either case, an engineer or analyst will use the collected data to develop an estimate of spillover savings for each project.

Sample Frame. One target for participant spillover research may be the most recent year's program participants who have been sampled for free ridership or process surveys. In the case where a stand-alone spillover study is being conducted, the sample frame may be broader and include those whose participation occurred during the time period of two prior program years.

Because evaluated spillover energy impacts associated with the sample are being extrapolated to the program population, it is important that the sample frame be limited to participating customers for which spillover may potentially be claimed.

Sample frames should be constructed in accordance with the following guidelines:

- Self-directing customers as defined by 220 ILCS 5/8-104(m) should be excluded from the sample frame for natural gas spillover.
- Customers of municipal electric utilities should be excluded from the sample frame for electric spillover.

Timing of Data Collection. Evaluators may either administer the participant spillover module as part of a comprehensive net-to-gross survey, or they may elect to implement it separately. A follow-up in-depth interview may also be conducted by an engineer or analyst to obtain additional details needed to quantify savings. Optimally, the spillover inquiry should be timed in order to allow sufficient time for spillover to occur; at a minimum, three months after the program-incented measure is installed. Projects installed up to two years after program participation occurred may be counted as spillover, provided it can be substantiated.

3.2.1.1.2 Approach for Identifying and Quantifying Spillover

Attribution Threshold Condition. Spillover cases are identified using a threshold approach, in which certain conditions must be met in order to qualify as a spillover measure. The threshold condition for spillover is based on responses to the following two survey questions:

1. How important was your experience in the <PROGRAM> in your decision to implement this measure, using a scale of 0 to 10, where 0 is not at all important and 10 is extremely important?
2. If you had not participated in the <PROGRAM>, how likely is it that your organization would still have implemented this measure, using a 0 to 10 scale, where 0 means you definitely WOULD NOT have implemented this measure and 10 means you definitely WOULD have implemented this measure?

The response to the first question cited above is “Measure Attribution Score 1,” and the response to the second question cited above is “Measure Attribution Score 2.” Spillover is considered to be attributable to the program if the “Spillover Score” is greater than 7.0. The “Spillover Score” is defined as follows:

$$\text{Spillover Score} = (\text{Measure Attribution Score 1} + (10 - \text{Measure Attribution Score 2}))/2$$

Spillover is considered to be attributable to the program if the average of the Measure Attribution Score 1 and (10 – Measure Attribution Score 2) exceeds 7.0; either the Measure Attribution Score 1 or (10 – Measure Attribution Score 2) could be below 7.0—as long as the average is greater than 7.0, the threshold is met.

If the average is greater than 7.0, 100% of the spillover savings referenced in the question are considered to be attributable to the program. If the average is not greater than 7.0, none of the spillover savings are considered to be attributable to the program.

Calculation of Spillover Measure Energy Savings. Energy savings of spillover measures shall be calculated in one of two ways.

1. Those addressed in the IL-TRM shall be calculated in accordance with the methods and algorithms specified in the IL-TRM, and shall reference the IL-TRM-defined time-of-sale or new construction baseline.
2. For measures not addressed in the IL-TRM, evaluators shall quantify savings using accepted industry-wide savings methods that conform to IPMVP or other industry protocols and documents.

Evaluators will make every effort to ensure that there is no double-counting of participant spillover energy savings across multiple sources of participant and nonparticipant spillover (such as participating customer and trade ally surveys) and will document that effort.

Measure implementation must have occurred within one year of the participant spillover study data collection effort in order to be countable as participant spillover.

For the purposes of accounting for spillover savings attributable to a program, spillover will only be quantified for measures implemented within the Program Administrator’s service territory.

3.2.1.1.3 Key Participant Spillover Survey Questions

The Participant Spillover question module is designed to be a general inquiry that seeks to: (1) assess whether additional energy efficiency improvements were implemented since the rebated project was completed; (2) confirm that these measures either had not received program incentives, or that there were no plans to submit them for program incentives in the future; (3) gather basic information about the additional energy efficiency measures (e.g., their type, size, quantities, and energy efficiency rating); and (4) establish program attribution.

The basic question structure is shown below. The measure-specific questions can be repeated in order to capture multiple measures. Note that there is considerable flexibility to tailor the questions to specific types of applications and programs.

1. Since your participation in the <PROGRAM>, did you implement any ADDITIONAL energy efficiency improvements at this facility or at your other facilities within <PROGRAM ADMINISTRATOR>’s service territory that did NOT receive incentives through <PROGRAM>?

2. What measures did you implement without an incentive?

MEASURE-SPECIFIC QUESTIONS [repeated for each spillover measure]³⁷

1. How important was your experience in the <PROGRAM> in your decision to implement this <MEASUREX>? Please use a scale of 0 to 10, where 0 is not at all important and 10 is extremely important.
2. Can you explain how your experience with the <PROGRAM> influenced your decision to install this additional high-efficiency measure?
3. If you had not participated in the <PROGRAM>, how likely is it that your organization would still have implemented <MEASURE>? Please use a 0 to 10, scale where 0 means you definitely WOULD NOT have implemented this measure and 10 means you definitely WOULD have implemented this measure.
4. How many of <MEASURE> did you install?
5. Questions to further define the measure (as applicable):
 - a. Type
 - b. Efficiency
 - c. Size
 - d. Other attributes
6. Can you briefly explain why you decided to install this energy efficiency measure on your own, rather than going through the <PROGRAM>?

3.2.1.1.4 Reporting of Results

Evaluators will report the following information relating to participant spillover data collection and analysis in annual EM&V reporting: 1) the number of participants surveyed; 2) the number of survey respondents reporting spillover; 3) the number of survey respondents who meet the spillover attribution threshold; 4) the number of respondents for which spillover savings were actually quantified; 5) the spillover savings for each project and overall; and 6) the spillover rate. The term (1-Free Ridership) is referred to as the Core NTGR.

The annual EM&V report should also describe the means by which the participant spillover rate is calculated. Two possible approaches are:

- (1) Add the participant spillover rate to each project's Core NTGR. The project-level NTGRs are then weighted by each project's ex ante or ex post (if available) gross savings as a share of the total. This savings-weighted NTGR can then be applied to the ex post gross savings of the participant population. If the sample is stratified, sampling weights must be applied before applying the NTGR to the ex post gross savings of the participant population.
- (2) Estimate program spillover effects by summing overall project-level spillover estimates for the sample and dividing this sum by the total ex ante or ex post (if available) gross savings for the sample to produce the participant spillover rate. This participant spillover rate can be added to the Core NTGR for the sample to yield the NTGR. If the sample is stratified, sampling weights must be applied before applying the NTGR to the ex post gross savings of the participant population.

In both cases, the participant spillover rate must be calculated at the project level for Option 1 or at the program level for Option 2, using the following formula.

³⁷ Example questions to gather engineering information to support the calculation of spillover savings may be accessed here: http://www.ilsag.info/il_ntg_methods.html

$$\text{Participant Spillover Rate} = \frac{ISO + OSO \text{ in sample}}{Ex Post Gross Impacts in sample}$$

Where:

ISO = Inside participant spillover

OSO = Outside participant spillover

3.2.1.2 Nonparticipant Spillover

The evaluation may perform research to measure nonparticipant spillover. Guidelines for the assessment of nonparticipant spillover may be added to this document in the future. Evaluators will make efforts to ensure that there is no double-counting of energy savings across multiple sources and will document those efforts.

3.3 Small Business Protocol

3.3.1 Free Ridership

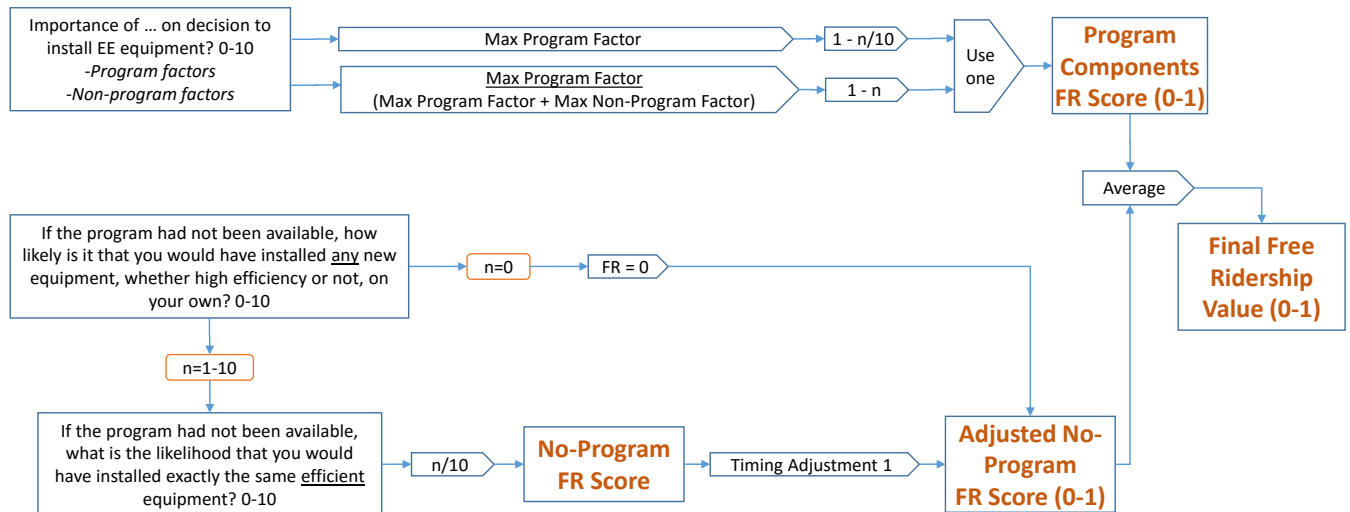
The FR algorithm for non-residential small business programs will follow the Core Non-Residential FR Protocol, with the following exceptions:

1. To reduce respondent burden, the Program Influence FR Score may be dropped from the Small Business FR algorithm. The influence of non-program factors will still be captured in the Program Components FR Score.
2. The counterfactual likelihood question (likelihood the participant would have installed the exact same energy efficiency equipment absent the program) may be preceded with a 0-10 scale question about the likelihood the participant would have installed any new equipment—either standard efficiency or high efficiency—on their own.
 - a. If the participant provides a likelihood response of 0, then the No-Program FR Score for that participant is set to 0.
 - b. If the participant provides a likelihood response of 1-10, then the participant is asked the same counterfactual questions (including the first timing question) as in the Core Non-Residential FR protocol.
3. To reduce respondent burden, the second question about timing (likelihood the participant would have installed the exact same energy efficiency equipment within 12 months) may be dropped. In this case, the only Deferred Free Ridership specification would be the one applying Timing Adjustment 1.

The diagram below, Figure 3-4, depicts the Small Business FR approach with exceptions 1 and 3 above implemented.

Figure 3-4. Small Business Free Ridership

(Program Components FR Score + (No-Program FR Score * Timing Adjustment 1)) / 2



Evaluators will calculate free ridership scores for small business projects as follows:

If Program Influence FR Score is dropped:

$$FR = \text{AVERAGE} ([\text{Program Components FR Score}], [\text{No-Program FR Score} * \text{Timing Adjustment 1}])$$

If Program Influence FR Score is included:

$$FR = \text{AVERAGE} ([\text{Program Components FR Score}], [\text{Program Influence FR Score}], [\text{No-Program FR Score} * \text{Timing Adjustment 1}])$$

Evaluators will develop estimates of free ridership based on the two Program Components FR Score specifications outlined in the Core Non-Residential FR protocol. Evaluators will select one of these for purposes of calculating the annual incremental energy savings for comparing to the legislated goal. Evaluators will present the results of both estimates of free ridership in EM&V reporting.

3.4 C&I New Construction Protocol

3.4.1 Free Ridership

The FR algorithm for non-residential new construction programs will follow the Core Non-Residential FR protocol, with the following exception:

- The concept of project timing and deferred free ridership is not applicable to new construction projects.³⁸ As a result, the various deferred free ridership specifications outlined in Figure 3-1, Figure 3-2, and Figure 3-3 will not be included in the free ridership estimation for new construction projects.

³⁸ New Construction programs intervene in the early phases of ongoing construction projects (i.e., after the decision to build has been made). As a result, participation in a New Construction program would not be expected to accelerate the construction of the new building.

Evaluators will calculate free ridership values for new construction projects as follows:

$$FR = \text{AVERAGE} ([\text{Program Components FR Score}], [\text{Program Influence FR Score}], [\text{No-Program FR Score}])$$

Evaluators will develop estimates of free ridership based on the two Program Components FR Score specifications outlined in the Core Non-Residential FR protocol and will select one of these for purposes of calculating the annual incremental energy savings for comparing to the legislated goal. Evaluators will present the results of both estimates of free ridership in EM&V reporting.

3.5 Study-Based Protocol

3.5.1 Free Ridership

The FR algorithm for non-residential study-based programs (See Figure 3-5) will follow the Core Non-Residential FR protocol, with the following exceptions:

- The counterfactual likelihood question (Q.4 in Figure 3-6 and Figure 3-7, below) will be preceded by five questions.³⁹
- Q.1 A 0-10 scale question about the likelihood that the participant would have conducted the study absent the program will be included.

At the measure-group level, the following should be included:

- Q.2a A yes/no question to determine if the participant performs regular maintenance on the equipment treated through the program
- Q.2b If the response to Q.2a is “yes,” a yes/no question to determine if the maintenance always includes the treatment provided through the program
- Q.3a A yes/no question to determine if the participant had prior awareness of the performance issues identified through the study
- Q.3b A 0-10 scale question about the participant’s level of familiarity with the recommended actions to rectify the performance issue.

The counterfactual likelihood question (Q.4 – likelihood the participant would have taken action absent the program) and the first counterfactual timing question (used to develop Timing Adjustment 1) will be asked at the measure-group level. Measure-group level responses will be aggregated to the project level, using savings-based weights.

There will be two options for developing the No-Program FR Score:

1. The measure-group level Adjusted No-Program FR Score will be developed following Algorithm 1 of the Core Non-Residential FR approach, using responses to the counterfactual likelihood question (Q.4) and Timing Adjustment 1.
2. The measure-group level No-Program FR Scores will be assigned, based on responses to Q.1, Q.2b, Q.3a, and Q.3b, as follows:
 - a. If Q.2b = Yes, then No-Program FR Score = 1. This assumes that if the participant performs regular maintenance on the treated equipment and that maintenance always includes the issue

³⁹ It should be noted that the question numbering in Figure 3-6 and Figure 3-7 is for reference purposes only; the additional questions do not have to immediately precede the counterfactual likelihood question.

addressed through the program, then the participant is a full free rider for that measure group for purposes of calculating the No-Program FR Score.

- b. If Q.3a = No and Q1 = 0 and Q.2b ≠ Yes, then No-Program FR Score = 0. This assumes that if the participant was not aware of the performance issue and had a zero likelihood of performing the study absent the program and their maintenance practices do not always include the issue addressed through the program, then the participant is not a free rider for that measure group for purposes of calculating the No-Program FR Score since they would not have found out about the issue absent the program.
- c. If Q.3b = 0 and Q1 = 0 and Q.2b ≠ Yes, then No-Program FR Score = 0. This assumes that if the participant had no familiarity with how to rectify the performance issue, had a zero likelihood of performing the study absent the program, and their maintenance practices do not always include the issue addressed through the program, then the participant is not a free rider for that measure group for purposes of calculating the No-Program FR Score since they would not have known how to address the issue absent the program.
- d. For all other combinations of responses to Q.1, Q.2b, Q.3a, and Q.3b, the measure-group level Adjusted No-Program FR Scores will be developed following Algorithm 1 of the Core FR approach, using responses to the counterfactual likelihood question (Q.4) and Timing Adjustment 1.

Figure 3-5. Study-Based Free Ridership—Overview

(Program Components FR Score + Program Influence FR Score + (No-Program FR Score * Timing Adjustment 1)) / 3

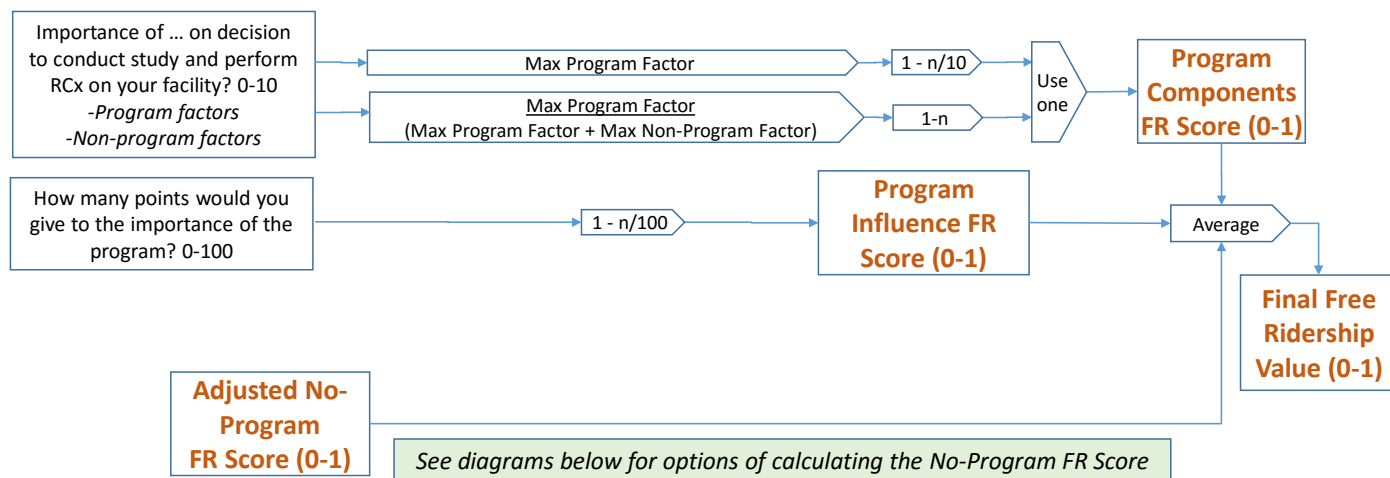


Figure 3-6. Study-Based Free Ridership—No-Program FR Score Option #1

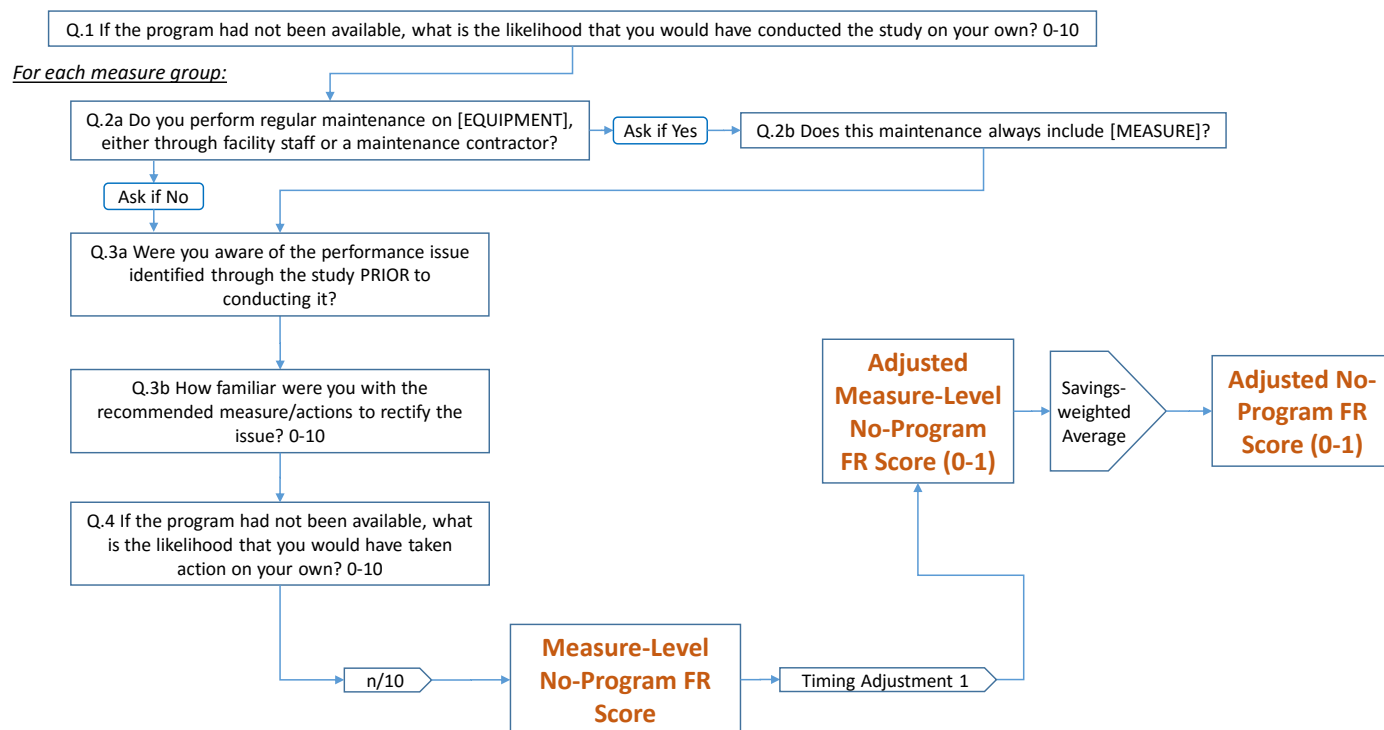
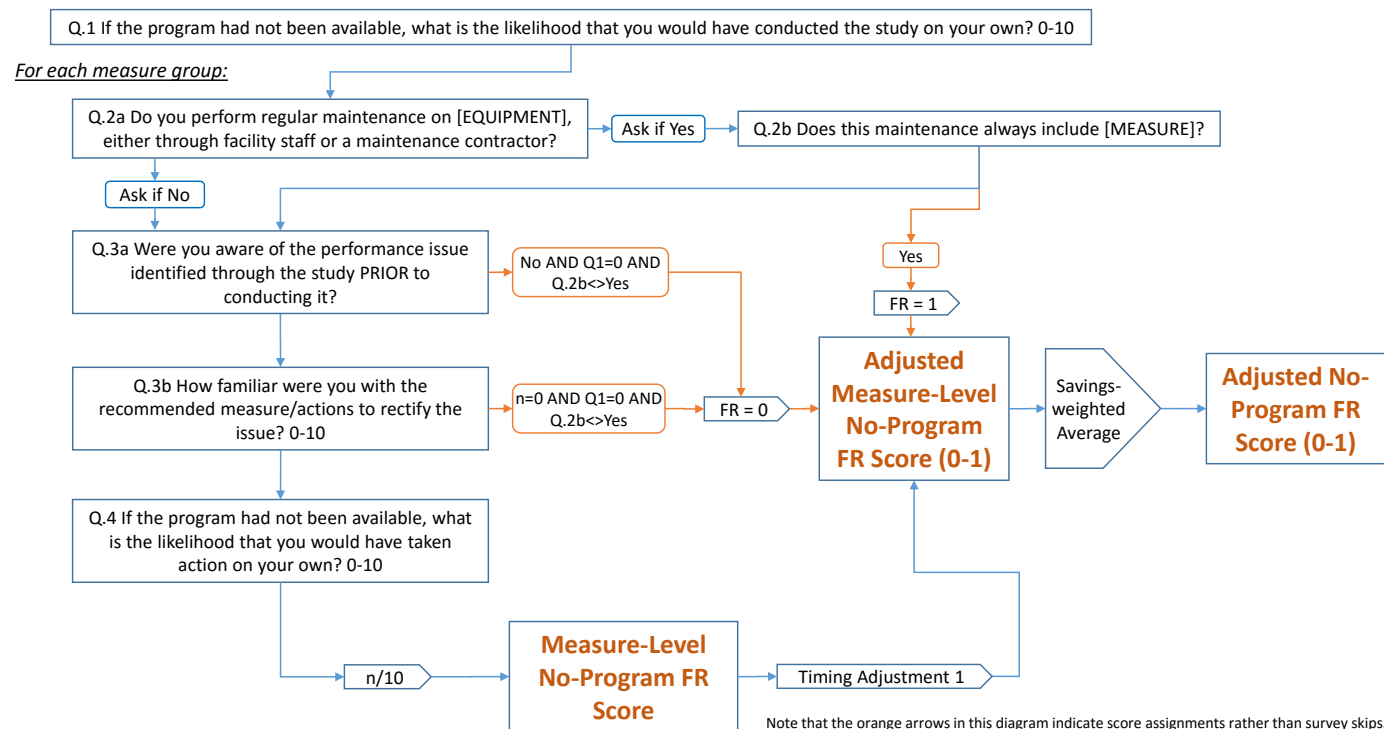


Figure 3-7. Study-Based Free Ridership—No-Program FR Score Option #2



Evaluators will calculate free ridership values for study-based programs as follows:

$$FR = \text{AVERAGE} ([\text{Program Components FR Score}], [\text{Program Influence FR Score}], [\text{No-Program FR Score} * \text{Timing Adjustment 1}])$$

Evaluators will develop estimates of free ridership based on the two No-Program FR Score options outlined above and the two Program Components FR Score specifications outlined in the Core Non-Residential FR protocol. Evaluators will select one of these for purposes of calculating the annual incremental energy savings for comparing to the legislated goal. Evaluators will present the results of both estimates of free ridership in EM&V reporting.

3.6 Training and Technical Assistance Protocol

This protocol is applicable to programs that provide training or technical assistance to encourage the adoption of energy efficiency measures in non-residential facilities, but do not provide financial incentives.

3.6.1 Program-Attributable Savings

Program-attributable savings are defined as energy savings resulting from the implementation of energy efficiency measures that were facilitated by assessment or technical assistance provided by the program and which did not receive a financial incentive through an Illinois energy efficiency program.

3.6.2 Research Methods

Data collection approach. An initial determination of program-attributable savings may be made based on self-reported findings from surveys of program participants. At a minimum, surveys collecting data pertaining to participant measure implementation will obtain general information on the specific measures installed and information substantiating their attribution to the program. Research on the specific characteristics of the energy-efficient equipment installed and the baseline and operating conditions needed to estimate savings may be done in one of two ways: 1) a detailed battery of measure specific questions may be administered as part of the initial survey; or 2) a separate in-depth follow-up interview may be conducted by the engineer or analyst responsible for the energy savings calculation. These collected data may be augmented by detailed facility and measure characteristics if provided by program staff.

3.6.2.1 Approach for Identifying and Quantifying Program-Attributable Savings

Attribution Threshold Condition. Program-attributable cases are identified using a threshold approach, in which certain conditions must be met in order to qualify as a program-attributable measure. The threshold condition for program attribution is based on responses to the following two survey questions:

1. How important was the information and/or assistance you received through the <PROGRAM NAME> in your decision to implement this measure, using a scale of 0 to 10, where 0 is NOT AT ALL IMPORTANT and 10 is EXTREMELY IMPORTANT?
2. If you had not received the information and/or assistance through the <PROGRAM>, how likely is it that your organization would still have implemented this measure, using a 0 to 10, scale where 0 means you definitely WOULD NOT have implemented this measure and 10 means you definitely WOULD have implemented this measure?

The response to the first question cited above is Measure Attribution Score 1, and the response to the second question cited above is Measure Attribution Score 2. An implemented measure is considered to be attributable to the program if the “Attribution Score” is greater than 7.0. The “Attribution Score” is defined as follows:

$$\text{Attribution Score} = (\text{Measure Attribution Score 1} + (10 - \text{Measure Attribution Score 2}))/2$$

The implemented measure is considered attributable to the program if the average of the Measure Attribution Score 1 and (10 – Measure Attribution Score 2) exceeds 7.0; either the Measure Attribution Score 1 or (10 – Measure Attribution Score 2) could be below 7.0—as long as the average is greater than 7.0, the threshold is met.

If the Attribution Score is greater than 7.0, 100% of the measure savings referenced in the question are considered to be attributable to the program. If the Attribution Score is equal to or less than 7.0, none of the measure savings are considered to be attributable to the program.

Calculation of Attributable Measure Energy Savings. Energy savings of program-attributable measures shall be calculated in one of two ways.

1. Those addressed in the IL-TRM shall be calculated in accordance with the methods and algorithms specified in the IL-TRM, and shall reference the appropriate IL-TRM-defined Time of Sale or New Construction baseline.
2. For measures not addressed in the IL-TRM, evaluators shall quantify savings using accepted industry-wide savings methods that conform to IPMVP and other industry protocols. All calculations should be done in a transparent manner that meets annual reporting requirements. The IPMVP protocols are to be used under the following situations: 1) the measure is non-standard and/or not directly addressed by the IL-TRM; 2) the expected energy savings are large; or 3) the evaluator believes that the precision of savings estimates would be substantially improved by its application.

Evaluators will make every effort to ensure that there is no double-counting of program-attributable energy savings across multiple sources of participant and nonparticipant spillover reporting (such as participating customer and trade ally surveys).

3.6.2.2 Reporting of Results

Evaluators will report the following information relating to data collection and analysis in annual EM&V reporting: 1) the number of participants surveyed; 2) the number of participants that implemented measures during the program year; 3) the number of survey respondents who meet the attribution threshold; 4) the number of respondents for which program-attributable savings were actually quantified; 5) the program-attributable savings for each project and overall; and 6) the total program-attributable savings.

Total program-attributable savings will be calculated in one of two ways. For programs that collect data on implemented projects and associated expected energy savings, total program attributable savings will be calculated by weighting the sum of verified program-attributable savings by the ratio of total program reported savings to sampled project savings. For programs that do not report projects implemented by program participants, total program-attributable savings will be calculated by weighting the sum of verified program-attributable savings by the ratio of total program participants to sampled program participants.

4 Residential and Low Income Sector Protocols

The table below lists Illinois residential programs and the NTG protocol applicable to each program.⁴⁰ If the design of a given program changes significantly, then it may mean that the NTG protocol listed for that program in this document is no longer appropriate. If that happens, the evaluator should follow the procedures outlined in Section 1.4: Diverging from the IL-NTG Methods.

Table 4-1. Residential and Low Income Programs

Program Administrator	Free Ridership Protocol	Program Name
Ameren Illinois	4.2 Appliance Recycling Protocol	Appliance Recycling
	4.3 Residential Upstream Lighting Protocol	Upstream Lighting
	4.4 Prescriptive Rebate (With No Audit) Protocol	Heating and Cooling
	4.5 Single-Family Home Energy Audit Protocol	All Electric Homes (Single Family)
		Home Performance with Energy Star
	4.6 Multifamily Protocol	All Electric Homes (Multifamily-Major Measures)
		Multifamily (In-Unit, Common Area and Major Measures)
	4.7 Energy Saving Kits and Elementary Education Protocol	Direct Mail Kits
		School Kits
	4.8 Residential New Construction Protocol	ENERGY STAR New Homes
ComEd	5.1 Behavioral Protocol	Behavior Modification
	NTG=1†	Moderate Income
	4.2 Appliance Recycling Protocol	Fridge and Freezer Recycling
	4.3 Residential Upstream Lighting Protocol	Lighting Discounts
	4.4 Prescriptive Rebate (With No Audit) Protocol	Appliance Rebates
		Heating and Cooling Rebates
		Weatherization Rebates
	4.5 Single-Family Home Energy Audit Protocol	Home Energy Assessments
	4.6 Multifamily Protocol	Elevate Energy Multifamily All Electric (In-Unit, Common Area and Major Measures)
		Residential Multifamily
	4.7 Energy Saving Kits and Elementary	Community-based CFL Distribution

⁴⁰ The “Free Ridership Protocol Name” in the second column of the table refers to the numbered sections in this document, e.g., “4.6 Multifamily Protocol.”

Program Administrator	Free Ridership Protocol	Program Name
	Education Protocol	Direct to Consumers Kits
		NTC Middle School Take Home Kits
		SuperSavers – Elementary Education Kits
	4.8 Residential New Construction Protocol	Residential New Construction
	5.1 Behavioral Protocol	Cub Energy Saver
		Home Energy Reports (Opower)
		Power Smart Reports
		Shelton Solutions Great Energy Stewards
	NTG=1†	Low-Income Kit Energy (LIKE)
	NTG=1†	Low-Income Multi-Family
Department of Commerce	4.7 Energy Saving Kits and Elementary Education Protocol	Energy Smart Schools
	NTG=1†	Affordable Housing Construction Program
	NTG=1†	Public Housing Authority Efficient Living Program
	NTG=1†	Residential Retrofit (Low Income)
Nicor	4.4 Prescriptive Rebate (With No Audit) Protocol	Home Energy Efficiency Rebates
	4.5 Single-Family Home Energy Audit Protocol	Home Energy Savings
	4.6 Multifamily Protocol	Multifamily Home Energy Savings
	4.7 Energy Saving Kits and Elementary Education Protocol	Elementary Energy Education
		Kits
	4.8 Residential New Construction Protocol	Residential New Construction
	5.1 Behavioral Protocol	Behavioral Energy Savings
Peoples Gas/ North Shore Gas	4.4 Prescriptive Rebate (With No Audit) Protocol	Home Energy Rebates
	4.5 Single-Family Home Energy Audit Protocol	Home Energy Jumpstart
	4.6 Multifamily Protocol	MF Custom
		MF Partner Trade Ally
		MF Prescriptive
		Multifamily (Direct Install)
	4.7 Energy Saving Kits and Elementary Education Protocol	Elementary Energy Education
	5.1 Behavioral Protocol	Home Energy Reports

Program Administrator	Free Ridership Protocol	Program Name
All	5.2 Code Compliance Protocol	Statewide Codes Collaborative

† The Uniform Methods Project notes that “most low-income programs are not subject to NTG analysis (that is, are deemed at 1.0).” In line with that common practice, there is general consensus among Illinois stakeholders that the Illinois low-income programs should not be subject to NTG analysis and thus the NTG ratios for low-income programs are effectively deemed at 1.0. See Violette and Rathbun (2014), Chapter 23: Estimating Net Savings: Common Practices. The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures, available electronically at <http://www.nrel.gov/docs/fy14osti/62678.pdf>, p. 50.

4.1 Residential Cross-Cutting Approaches

The approaches in this section can apply to more than one program type but do not supersede program-specific approaches presented in later sections.

4.1.1 Survey Design Issues

Free ridership questions should be asked near the beginning of a participant survey, before asking satisfaction questions. This should prevent participants from confusing free ridership questions with the satisfaction questions, which would result in an artificially high free ridership score, especially for participants highly satisfied with the measures.

4.1.2 Participant Spillover

Effective program marketing and outreach generates program participation and increases general energy efficiency awareness among customers. Spillover can be calculated using participant survey questions, which ask participants about energy-savings actions they have taken on their own since participating in the program. Questions should be sufficiently specific to ensure energy savings associated with spillover can be reasonably well-quantified. These may include questions about measure types or measures installed, quantities, and efficiency levels. When program implementers provide recommendations to participants and can provide data on the types of recommendations made to specific participants, evaluations should attempt to determine whether participants took the recommended actions outside of the program at sites within the program administrator’s service territory; if so, savings from those recommended actions should be attributed to the program.

To reduce the respondent’s burden, the survey should first ask participants about the influence the program had on their decision to take additional energy-saving actions on their own. In particular, the evaluation team should ask two close-ended questions to determine program influence on spillover actions. The two required questions, preceded by an optional open-ended warm-up question, are:

- OPTIONAL: Did the program influence you in any way to make these additional improvements?
- 1. How important was your participation in the <PROGRAM ADMINISTRATOR’S> program on your decision to make additional energy efficiency improvements on your own? [Scale from 0-10 where 0 is “not at all important” and 10 is “extremely important”]
- 2. If you had not participated in the <PROGRAM ADMINISTRATOR’S> program, how likely is it that you would still have implemented this measure, using a 0 to 10, scale where 0 means you definitely WOULD NOT have implemented this measure and 10 means you definitely WOULD have implemented this measure?

The response to the first required question cited above is “Measure Attribution Score 1,” and the response to the second required question cited above is “Measure Attribution Score 2.” The specific measures referenced in the question are considered to be attributable to the program if the “Spillover Score” is greater than 7.0:

$$\text{Spillover Score} = (\text{Measure Attribution Score 1} + (10 - \text{Measure Attribution Score 2}))/2 > 7.0$$

If these conditions are met, the evaluator determines that the specific measures referenced in the question are attributable to the program; otherwise, the evaluator determines that the specific measures referenced in the question are not attributable to the program. The attribution criterion represents a threshold approach, in which energy impacts associated with measures implemented by program participants outside the program are either 100% program-attributable or 0% program-attributable.

For each measure mentioned, customers will be asked how they know the measure is more efficient than other models. If the respondent can identify the measure as ENERGY STAR or name an efficiency level that the evaluator confirms as being above the minimum federal standard, or if they identify a technology that the evaluator can confirm is above the minimum federal standard, it will count towards Participant Spillover.

Finally, depending on the measure type cited by the customer, follow-up questions should ask customers to provide reasonable information to allow the evaluator to estimate the amount of savings using IL-TRM protocols, such as quantity of appliances or the location and amount of insulation.

To calculate the spillover energy and demand savings for these actions, the appropriate version of the IL-TRM should be used. To develop the spillover rate, the total energy and demand impacts from the sampled participants who installed additional measures due to participation in the program are summed, and then this sum is divided by the total ex post sample energy and demand impacts:

$$\text{Participant Spillover Rate (PSO)} = \frac{\text{Sum of Energy or Demand from Additional Measures Installed}}{\text{Sample Ex Post Gross Energy or Demand Impacts}}$$

The equation used to adjust the Core NTGR based on participant spillover is as follows:

$$\text{NTGR} = (1 - \text{FR} + \text{PSO})$$

4.1.2.1 Data Collection

Respondents should be drawn from a random sample of current or up to one year of previous program participants. Regardless of the participation year, spillover should be measured within the last 12 months (from the survey date), but after previous participation; the tracking database should supply this information.

4.1.2.2 Data Analysis

The following four steps calculate spillover:

1. Calculate total spillover savings for each participant installing an efficient measure not rebated through the program where the Spillover Score is greater than 7.0:

$$\text{Measure Spillover} = \text{Measure Savings} * \text{Number of Units}$$

2. Total savings associated with each program participant to calculate overall participant spillover savings.

3. Spillover Percentage Estimate =
$$\frac{\sum \text{Sample Spillover kWh Savings}}{\text{Sample Evaluated Program kWh Savings}}$$

4.1.3 Nonparticipant Spillover Measured Through Trade Allies

In addition to participant free ridership and spillover, residential programs may create nonparticipant spillover (NPSO) through trade allies exposed to the program but not actually facilitating program participation. Rather, they promote and stock higher-efficiency equipment due to the program.⁴¹ NPSO caused by trade allies can be determined by surveying three groups of trade allies:

- Participating trade allies that do not submit rebates or otherwise act as program agents on behalf of their customers. For this group, care should be taken to ensure spillover is not double-counted with program sales.
- “Drop out” trade allies, who participated in the program previously but have not participated in the past 12 months.
- True nonparticipating trade allies that report they were aware of the program but had never participated.

Surveys ask nonparticipating trade allies if the program influenced their sales of high-efficiency equipment to nonparticipating customers and to quantify the program’s impact on their high-efficiency sales. The general questions take the following form:

- Q.1: How many <measures> did you sell in <period>?
- Q.2: How many of them were <efficiency level> or higher?
- Q.3: Had the <program> not existed, how many <measures> do you think you would have sold?

Evaluators should ensure that trade allies receive sufficient time to collect specific data and not rely on “guesses” to respond. Additional questions should be included to document how the program influenced sales of additional measures and to determine whether the responses to Question 3 should be included as NPSO. Responses should also clarify whether sales counts are specific to the utility service territory in question.

The following steps calculate the program’s nonparticipant trade ally spillover percentage:

1. Compute the difference between the total reported number of high-efficiency units sold and the total that would have been sold in the program’s absence to obtain the total number of spillover units for that trade ally.
2. Multiply the total net number of spillover units of each measure sold by each surveyed trade ally by the average gross unit savings for each measure type.
3. Sum the result for each contractor from the previous step, and weight the results by the ratio of the population of non-active trade allies to the sample to compute the total spillover energy over the program period.
4. Divide the spillover energy savings by program gross savings.

4.1.4 Nonparticipant Spillover Measured from Customers

The evaluation may perform research on the general population to measure nonparticipant spillover (NPSO). If so, care should be taken to ensure spillover is not double-counted with a trade-ally approach. The basic method uses a general population survey. An enhanced method for estimating spillover would use site visits to verify and gather information to estimate spillover savings in cases where significant spillover is reported by nonparticipants.

⁴¹ NPSO also can arise from nonparticipating customers as a direct result of general energy efficiency education and promotion efforts. A separate protocol addresses such NPSO. Care should be taken to ensure the different approaches do not double-count NPSO.

4.1.4.1 Basic Method

4.1.4.1.1 Sampling

As spillover may be rare in the general, nonparticipating population, determining spillover will likely require a large sample of customers who have not participated in any energy efficiency programs, including a behavioral program, within the past three years. Customers will be removed from the sample frame if their account numbers can be cross-referenced against a list of program participants from the previous three years. The survey should target household members responsible for paying utility bills.

4.1.4.1.2 Measures-Specific Questions

Depending on the spillover measure type reported by the customer, follow-up questions should be included to gather sufficient information to reasonably assess the saving amount by applying the IL-TRM, understanding that assumptions must be made if IL-TRM inputs cannot be easily supplied by the participant. Such assumptions should be conservative, or, if not conservative, reasons for deviating from the conservative application should be documented. Measures that cannot be reasonably quantified within available evaluation budgets should be excluded from spillover calculations.

For measures included in the IL-TRM, savings will be assessed using the IL-TRM algorithms. Baselines for measures not in the IL-TRM will be assessed based on appliance standards and building codes, if applicable, and, if not, through engineering judgements of existing or market conditions. Engineering assumptions and analysis by the evaluator will be applied for measures not included in the IL-TRM. Key assumptions should be documented in the report.

4.1.4.2 Attribution Approach

To receive credit for energy savings, the nonparticipant must fit the following criteria: (1) be familiar with the Program Administrator's energy efficiency campaign (e.g., ActOnEnergy for Ameren); and (2) indicate that some aspect of the Program Administrator's energy efficiency programs motivated their purchasing decisions. Influence will be measured on a scale of 0 to 10, where 10 is very influential and 0 is not at all influential. Savings attribution requires a Spillover Score of greater than 7.0.

Survey respondents will be asked series of questions following the logic shown in Figure 4-1. First, the customer will indicate whether they know about their Program Administrator's energy efficiency programs and/or marketing messages. Customers aware of the services will be asked a follow-up question (whether they have participated in a program in the past three years) to confirm their household qualifies as a true nonparticipant. If confirmed as a nonparticipant, the customer will be asked if they or anyone in their household made an energy efficiency improvement within the last year, and if so, what improvements they made. Responses to these questions will generate a list of potential spillover measures (shown at point "[A]" in Figure 4-1). If attribution is established as described further below, for each measure mentioned, customers will be asked how they know the measure is more efficient than other models. If the respondent can identify the measure as ENERGY STAR or name an efficiency level that the evaluator confirms as being above the minimum federal standard, or if they identify a technology that the evaluator can confirm is above the minimum federal standard, it will count towards NPSO.

To assess attribution for each spillover measure mentioned, the customer will be asked questions to be scored in two areas. Spillover may be program-attributable for those measures for which self-report data meet the following threshold condition:

$$\text{Spillover Score} = (\text{Attribution Score 1} + (10 - \text{Attribution Score 2}))/2 > 7.0$$

4.1.4.2.1 Attribution Score 1

The first score, "Attribution Score 1," measures the influence level (on a scale of 0 to 10, where 10 is very influential and 0 is not at all influential) their Program Administrator had on the decision to purchase the measure.

Influence can derive from the following:

1. General information about energy efficiency provided by the Program Administrator.
2. Word-of-mouth from people installing energy-efficient equipment and receiving a rebate from the Program Administrator.
3. Personal experience with a previous Program Administrator rebate program (longer than three years ago).
4. Information from a contractor or retailer communicating about the Program Administrator's programs.

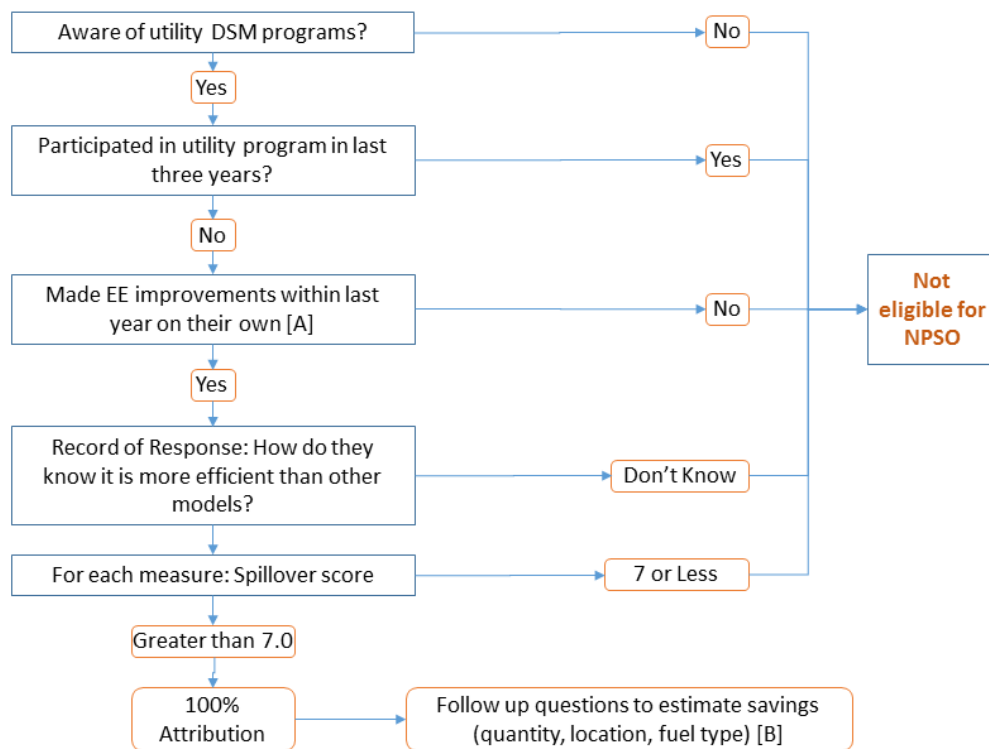
4.1.4.2.2 Attribution Score 2

The second score, "Attribution Score 2," comes from the customer's response to a single question to assess the counterfactual, asking about the likelihood (on a scale of 0 to 10, where 10 is highly likely and 0 is not at all likely) that the customer would have installed the measure had they not been influenced by the program.

The Spillover Score is then the average of the Attribution Score 1 and (10 – Attribution Score 2). If that Spillover Score is greater than 7.0, 100% of the savings are attributed to the Program Administrator for that measure.

Finally, depending on the measure type cited by the customer, follow-up questions will gather information to enable an estimate of savings (shown in the figure as [B]), such as quantity of appliances or the location of insulation.

Figure 4-1. NPSO Question Logic



4.1.4.3 Scoring

Survey respondents' answers to the NPSO questions will determine total energy and demand savings attributed to the program. Table 4-2 lists NPSO measures under column A, the Spillover Score under column B, the estimated measure savings under column C, the percentage of allocated savings under column D, and the total allocated

savings under column E. Column F shows the calculated average energy savings per spillover measure, determined by dividing the total allocated savings (the sum of column E) by the number of surveyed nonparticipating customers. The table shows how kWh NPSO savings would be calculated; calculations of therm or demand savings would be accomplished in the same manner.

Table 4-2. Estimation of Respondents' NPSO Savings

A	B	C	D	E	F
Spillover Measure	Spillover Score	Measure Savings (kWh)	Allocated Savings	Total kWh Savings	Average kWh Per Surveyed Customer
Measure1	Scale of 0 to 10	Savings1	100% if [B] > 7.0	[C] x [D]	N/A
Measure2	Scale of 0 to 10	Savings2		[C] x [D]	
MeasureN	Scale of 0 to 10	SavingsN	0% if [B] ≤ 7.0	[C] x [D]	
				Sum of column E = Total kWh Savings	Total kWh Savings ÷ Number of Completed Surveys

Table 4-3 shows the process for estimating total NPSO generated by the Program Administrator during the program year (for electric savings). The savings attributed from the survey population will be extrapolated to the nonparticipating residential customer population to determine the overall NPSO savings. Then NPSO energy savings will be converted into a percentage using the total evaluated electric savings for the program year. A similar process would apply for calculating therm or demand NPSO.

Table 4-3. Calculation of Total NPSO Generated

Variable	Description	Source/Calculation
F	Average kWh Energy Savings per Surveyed Customer	Survey data and impact evaluation
J	Total Nonparticipating Residential Population	Customer database
K	NPSO MWh Energy Savings Extrapolated to Nonparticipating Population	$[F \times J] \div 1,000 \text{ kWh/MWh}$
S	Total Evaluated MWh Savings	Residential Portfolio Savings
G	NPSO Spillover Rate	$K \div S$

4.2 Appliance Recycling Protocol

Appliance recycling programs (ARPs) typically offer some mix of incentives and free pickups for the removal of old but operable refrigerators, freezers, or room air conditioners. These programs encourage consumers to undertake the following:

- Discontinue use of secondary or inefficient appliances;
- Relinquish appliances previously used as primary units upon their replacement (rather than keeping the old appliance as a secondary unit); and
- Prevent the continued use of old appliances in other households through direct transfers (i.e., giving it away or selling it) or indirect transfers (resale in the used appliance market).

As the program theory and logic for appliance recycling differ significantly from standard “downstream” incentive

programs (which typically offer rebates for purchases of efficient products), the free ridership estimation approach also significantly differs.

The basic and enhanced methods are described next.

4.2.1 Basic Method

4.2.1.1 Free Ridership

Free ridership is based on participants' anticipated plans had the program not been available, thus classifying a free rider as a participant who would have removed the unit from service regardless of the program.

Estimating net savings for ARPs should adopt a multistep process to segment participants into different groups, each with specific attributable savings.

In general, independent of program intervention, participating appliances would have been subject to one of the following options:

1. The appliance would have been kept by the participating household.
2. The appliance would have been discarded in a way that transfers the unit to another customer for continued use.
3. The appliance would have been discarded in a way that would have permanently removed the unit from service.

Only Option 3 constitutes free ridership (the proportion of units that would have been taken off the grid absent the program). Options 1 and 2 both indicate non-free riders. However, these respondents need to be further classified to account for potential induced replacement and secondary market impacts, both described below.

4.2.1.1.1 Data Collection

A participant survey—drawn from a random sample of participants—will serve as the primary source of data collected for estimating NTG for the ARP. To determine the percentage of participants in each of the three options, evaluators will begin by asking surveyed participants about the likely fate of their recycled appliance had it not been decommissioned through the program. Responses provided by participants generally can be categorized as follows:

1. Kept the appliance.
2. Sold the appliance to a private party (either an acquaintance or through a posted advertisement).
3. Sold or gave the appliance to a used-appliance dealer.
4. Gave the appliance to a private party, such as a friend or neighbor.
5. Gave the appliance to a charity organization, such as Goodwill Industries or a church.
6. Had the appliance removed by the dealer from whom the new or replacement appliance was obtained.
7. Hauled the appliance to a landfill or recycling center.
8. Hired someone else to haul the appliance away for junking, dumping, or recycling.

Additional, follow-up questions will be included to validate the viability of all responses.

Next, evaluators will assess whether each participant's final response indicates free ridership:

- Some final responses clearly indicate free ridership, such as: "I would have taken it to the landfill or recycling center myself."
- Other responses clearly indicate no free ridership, as when the appliance would have remained active within the participating home ("I would have kept it and continued to use it") or used elsewhere within

the Program Administrator's service territory ("I would have given it to a family member, neighbor, or friend to use").

If the respondent planned to have the unit picked up by the retailer and the retailer would likely resell the unit in the secondary market, they are not a free rider. Absent retailer survey primary research described in the Enhanced Options below, the evaluators will utilize data from the most recent research conducted of the ComEd program to determine the proportion of free riders unless another metric is mutually agreed upon by the evaluators.⁴²

Secondary Market Impacts

In the event that the unit would have been transferred to another household (Option 2 above), the question then becomes what purchasing decisions are made by the would-be acquirers of participating units now that these units are unavailable. Such would-be acquirers could:

1. Not purchase/acquire another unit.
2. Purchase/acquire another used unit.

Adjustments to savings based on these factors are referred to as the program's secondary market impacts.

If it is determined that the participant would have directly or indirectly (through a market actor) transferred the unit to another customer on the grid, the next question addresses what that potential acquirer did because that unit was unavailable. There are three possibilities:

- A. **None of the would-be acquirers would find another unit.** That is, program participation would result in a one-for-one reduction in the total number of appliances operating on the grid. In this case, the total energy consumption of avoided transfers (participating appliances that otherwise would have been used by another customer) should be credited as savings to the program. This position is consistent with the theory that participating appliances are essentially convenience goods for would-be acquirers. (That is, the potential acquirer would have accepted the appliance had it been readily available, but because the appliance was not a necessity, the potential acquirer would not seek out an alternate unit.)
- B. **All of the would-be acquirers would find another unit.** Thus, program participation has no effect on the total number of appliances operating on the grid. This position is consistent with the notion that participating appliances are necessities and that customers will always seek alternative units when participating appliances are unavailable.
- C. **Some of the would-be acquirers would find another unit, while others would not.** This possibility reflects the awareness that some acquirers were in the market for an appliance and would acquire another unit, while others were not (and would only have taken the unit opportunistically).

The evaluators will assume Possibility C unless primary research within a Program Administrator's service territory to assess the secondary appliance market is undertaken as described in the Enhanced Options below. Specifically, evaluators will assume that half (0.5, the midpoint of Possibilities A and B) of the would-be acquirers of avoided transfers found an alternate unit.

Once the proportion of would-be acquirers who are assumed to find alternate units is determined, the next question is whether the alternate unit was likely to be another used appliance (similar to those recycled through the program) or, with fewer used appliances presumably available in the market due to program activity, would the customer acquire a new standard-efficiency unit instead.

⁴² Note that such retailer interviews are being conducted annually for the ComEd ARP evaluation, and answers are used directly in the calculation of the NTG ratio in cases where: (1) the respondent planned to have the unit picked up by the retailer; and (2) the retailer was interviewed.

4.2.1.2 Induced Replacement

If, however, the unit would have been kept by the participating household, the next question is whether the appliance was replaced and, if so, whether the household would have replaced the appliance regardless of the program.

The purchase of a refrigerator in conjunction with program participation does not necessarily indicate induced replacement. (The refrigerator market is continuously replacing older refrigerators with new units, independent of any programmatic effects.) However, if a customer would have not purchased the replacement unit (put another appliance on the grid) in the absence of the program, the net program savings should reflect this fact. This is, in effect, akin to negative spillover and will be used to adjust net program savings downward.

Estimating the proportion of households induced to replace their appliance should be done through participant surveys. As an example, participants could be asked, “Would you have purchased your replacement refrigerator if the recycling program had not been offered?”

Because an incentive ranging from \$35 to \$50 is unlikely to be sufficient motivation for purchasing an otherwise-unplanned replacement unit (which can cost \$500 to \$2,000), it is critical that evaluators include a follow-up question. That question should confirm the participants’ assertions that the program alone caused them to replace their refrigerator. For example, participants could be asked, “Let me be sure I understand correctly. Are you saying that you chose to purchase a new appliance because of the appliance recycling program, or are you saying that you would have purchased the new appliance regardless of the program?”

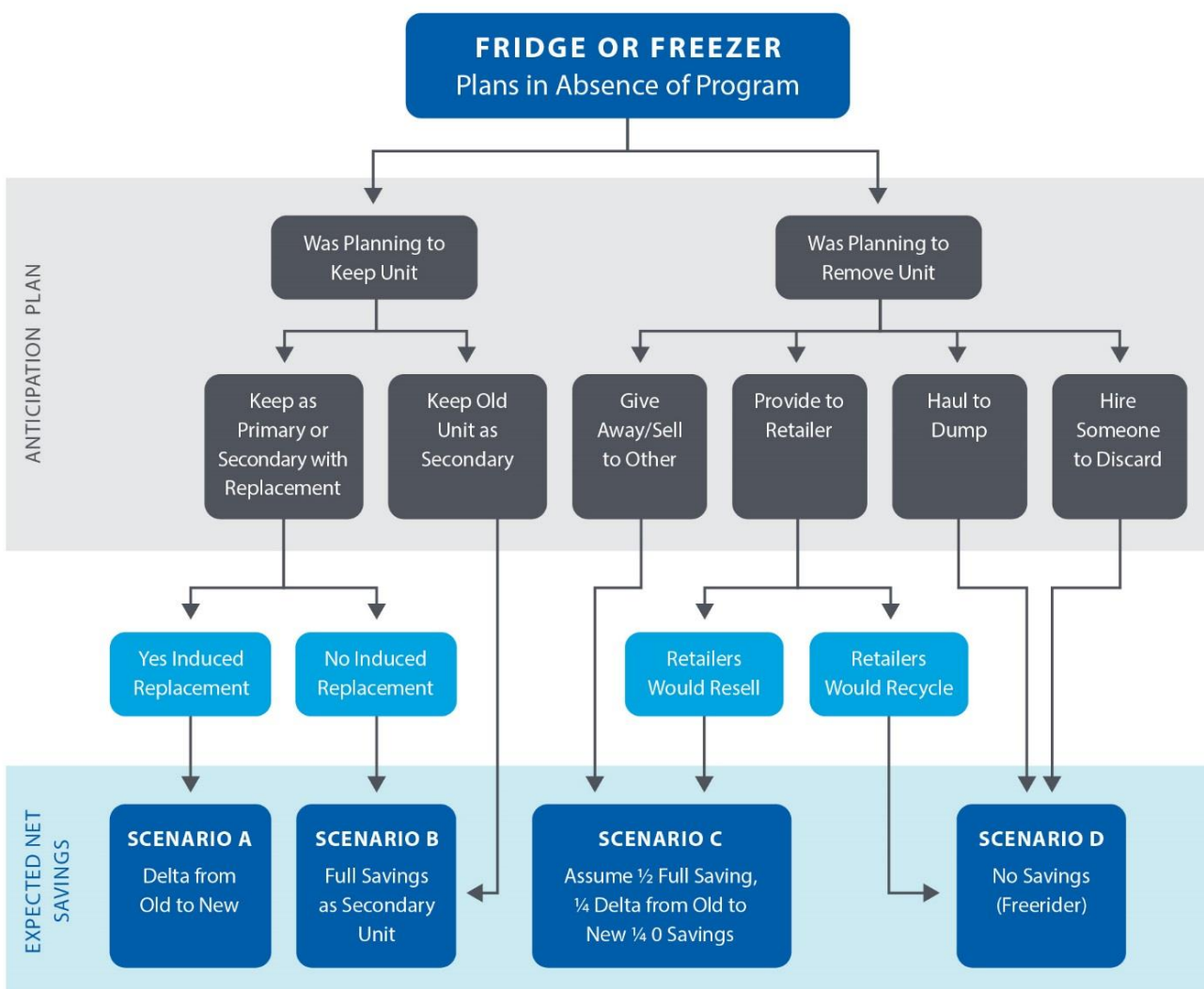
When assessing participant survey responses to calculate induced replacement, evaluators will consider the appliance recycled through the program as well as the participant’s stated intentions in the absence of the program. For example, if customers indicate they would have discarded their primary refrigerator independent of the program, it is not possible that the replacement was induced (because it is extremely unlikely the participant would live without a primary refrigerator). Induced replacement is a viable response for all other usage types and stated intention combinations.

As one might expect, previous evaluations have shown the number of induced replacements to be considerably smaller than the number of naturally occurring replacements unrelated to the program. Once the number of induced replacements is determined, this information is combined with the energy consumption replacement appliance to determine the total energy consumption induced by the program (on a per-unit basis).

4.2.1.3 Integrating Free Ridership, Secondary Market Impacts, and Induced Replacement

The flow chart shown in Figure 4-2Figure illustrates how net savings will be derived for an ARP. As shown, below, expected savings fall into four different scenarios.

Figure 4-2. Appliance Retirement Scenarios



Source: Adapted from the Pennsylvania Statewide Evaluator Common Approach for Measuring Net Savings for Appliance Retirement Programs, Guidance Memo-026, March 14, 2014.

4.2.1.4 Scoring Algorithm

Net savings will be assigned individually to each respondent, based on responses provided to the questions discussed above. Net savings will be averaged across all respondents to calculate program-level net savings. The following equation will be used:

$$FR = (\text{free ridership and secondary market impacts \%} - \text{induced replacement \%})$$

Table 4-4 demonstrates the proportion of a sample population classified into each of the eight potential (Tertiary Classification) categories and the resulting weighted net savings.

Table 4-4. Net Savings Example for a Sample Population*

Table 1: kWh Savings Example for a Sample Population						
Primary Classification	Secondary Classification	Tertiary Classification	Population (%)	UEC (kWh) w/out Program	UEC (kWh) w/ Program	kWh Savings
Would have kept unit	Scenario A: Kept but Induced Replacement	Non-ES unit	3%	1,026	520	506
		ES unit	2%	1,026	404	622
	Scenario B: Kept but NO Induced Replacement	N/A	25%	1,026	0	1,026
Would have removed unit	Scenario C1: Transferred No Induced Replacement	N/A	30%	1,026	520	506
	Scenario C2: Transferred With Induced Replacement	Non-ES unit	3.5%	1,026	520	506
		ES unit	3.5%	1,026	404	622
	Scenario D: Removed from Service	Recycled/Destroyed	20%	0	0	0
		Retailer would Recycle	13%	0	0	0
	Net Savings (kWh)					

*The percent values presented in this table serve only as examples; actual research should be conducted to determine the percentage of units falling into each of these categories. Note that UEC (Unit Energy Consumption) values presented in the table represent example values, factoring in part-use.

4.2.2 Enhanced Method

Results can be enhanced by including three additional research efforts. The basic method has defaults where primary research on enhanced approaches cannot be performed:

1. A retailer survey, to determine the quantity and/or proportion of units returned to a retailer and that the retailer would deconstruct or recycle. Through this survey, one would determine a retailer's criteria for reselling used units vs. deconstructing them, based on unit age and condition. Results from the survey and analysis would be used to determine the proportion of those who would have returned an old appliance

to the retailer that should be included in Scenario D (free riders). This research was conducted for ComEd in EPY6 evaluation and those results were applied to Ameren.

2. An appliance market assessment study to determine the size of the secondary appliance market and whether removal of participating units from the market would cause an otherwise would-be receiver to purchase an alternative used or new unit. Savings attributable to these participants are the most difficult to estimate, as the scenario attempts to estimate what the prospective buyer of a used appliance would do in the absence of finding a program-recycled unit in the marketplace (i.e., the program took the unit off the grid, so the prospective purchaser faced, in theory, a smaller supply of used appliances). It is difficult to answer this question with certainty, absent Program Administrator-specific information regarding the change in the total number of appliances (overall and used appliances specifically) that were active before and after program implementation. In some cases outside of Illinois, evaluators have conducted in-depth market research to estimate both the program's impact on the secondary market and the appropriate attribution of savings for this scenario. Although these studies are imperfect, they can provide Program Administrator-specific information related to the program's net energy impact. Where feasible, evaluators and utilities should design and implement such an approach. Unfortunately, this type of research tends to be cost-prohibitive, or the necessary data may simply be unavailable.
3. However, it is possible to estimate through nonparticipant surveys which of the disposal responses given by nonparticipants were most likely to have been to an opportunistic would-be-acquirer. Transfers that would most likely have been opportunistic are determined primarily based on the cost to the recipient. If the appliance was sold or transferred to a retailer, there would have been a cost to the recipient of that appliance. If the recipient was willing to pay for the appliance or was willing to exert the effort to visit a retail location, this suggests the recipient was actively seeking an appliance. However, if the unit were given away for free, there was little cost to the recipient and it is a reasonable proxy for the proportion of opportunistic acquirers. This proportion would replace the 50% default assumption (scenario C in Figure 4-2Figure) of would-be-acquirers that would or would not find an alternate unit.
4. A nonparticipant survey can be used to assess how nonparticipants acquire and dispose of used units. As nonparticipants do not have the same perceived response bias as participants, they can help offset some of this potential bias in estimating the true proportion of the population that would have recycled their units in program's absence. The evaluators will average the results of the nonparticipant survey with the participant survey if the nonparticipant survey is of sufficient sample size. Otherwise, results may be used for a qualitative characterization of potential bias. Though recommended, use of a nonparticipant survey need not be required, given budget and time considerations. A nonparticipant survey was completed as part of ComEd's EPY6 evaluation and used qualitatively to validate participant results.

4.3 Residential Upstream Lighting Protocol

The Illinois Residential Upstream Lighting programs to date have provided discounts on efficient lighting through retailers at the point of purchase. Such programs often remain transparent to customers purchasing incentivized lighting. Program administrators also do not know the identity of most customers purchasing the program-discounted lighting; so these customers cannot easily be contacted once they leave the store for a traditional self-report NTG evaluation survey (i.e., an after-the-fact, direct solicitation of customers regarding what they would have done in the program's absence). Similar surveys can be conducted with customers within program retailers after they have made their lighting purchasing decision but before they leave the store. For programs such as this, in store customer surveys are preferable to the traditional self-report telephone surveys that ask customers to recall their past light bulb purchases. Light bulbs are a small and relatively insignificant purchase for most people, thus the recall bias could be substantial.

Further, as upstream programs work with multiple market actors and can include wide-reaching marketing campaigns promoting energy efficiency to the general public, they tend to stimulate spillover and "market effects." As a result, estimating NTG for upstream residential lighting programs can be challenging. Multiple methods exist, each with their own strengths and weaknesses.

Ameren and ComEd implement their residential lighting programs comparably, and the evaluation teams have used a consistent primary NTG evaluation method. This section details the consensus NTG methodology, which has been used multiple times for both ComEd and Ameren and is considered the most well-vetted and defensible NTG method that has been successfully used in Illinois.

For EPY5 and EPY6, Ameren and ComEd used a customer self-report methodology to estimate NTG for their upstream residential lighting programs.⁴³ Customer self-report data in this method are collected during surveys conducted within program retailers with customers purchasing program bulbs (i.e., in-store intercept surveys). This method separately estimates free ridership, participant spillover, and nonparticipant spillover. Details follow on the primary data collection and scoring algorithms.

4.3.1 Basic Method

4.3.1.1 Free Ridership

Free ridership for this program is calculated as the proportion of program bulbs that would have been purchased if the program did not exist. Three alternative scenarios could occur:

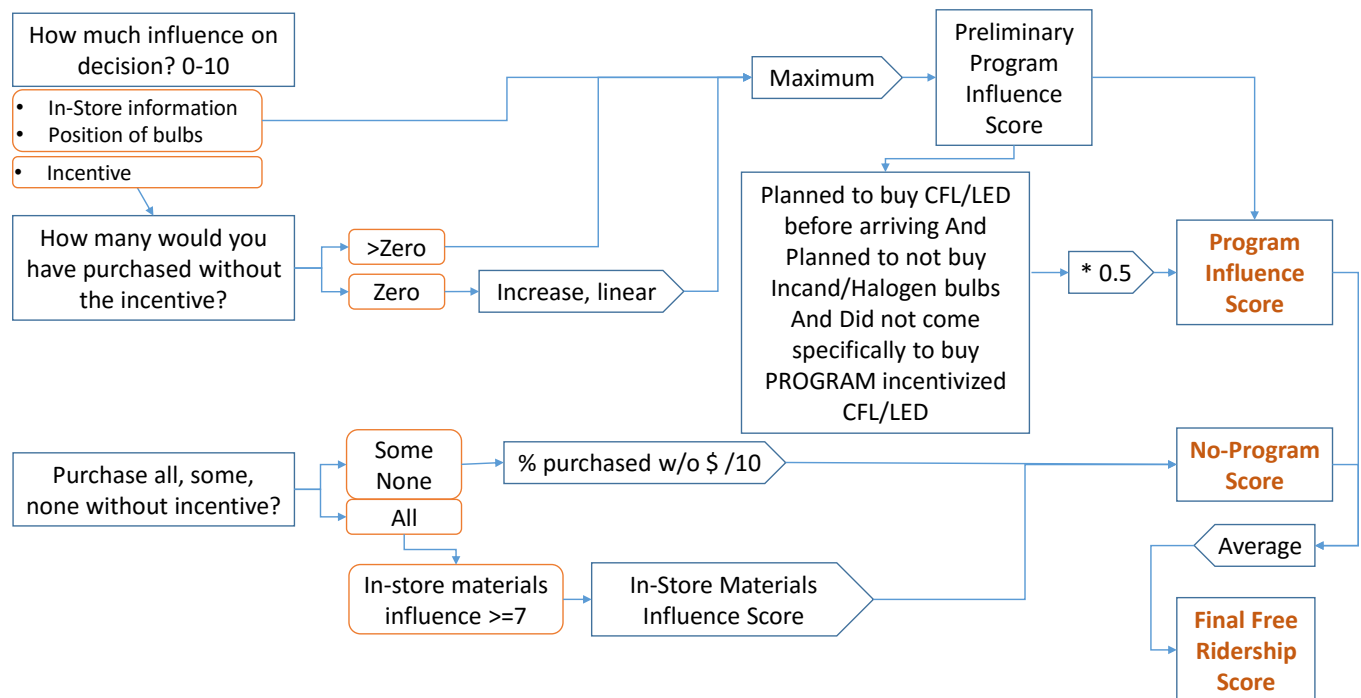
1. Full Free Rider: The customer would have purchased the same quantity of efficient bulbs (CFLs or LEDs) in the program's absence.
2. Partial Free Rider: The customer would have purchased fewer efficient bulbs (CFLs or LEDs) in the program's absence.
3. Non-Free Rider: The customer would have not purchased any efficient bulbs (CFLs or LEDs) in the program's absence.

Free ridership is calculated as the average of two distinct scores: a Program Influence Score and a No-Program score. These scores are defined as follows:

1. The Program Influence Score captures the maximum level of program influence, reported by a survey respondent, of the residential lighting program on their decisions to purchase program bulbs on the day of the survey. This program influence can take a number of forms, such as: the monetary incentive provided to decrease the cost of high-efficiency bulbs; program-sponsored educational materials that explain the benefits of efficient lighting; in-store product placement of efficient bulbs; and program bulb recommendations provided by retail store personnel.
2. The No-Program Score is used to estimate how many program bulbs a survey respondent would have purchased in the absence of the residential lighting program.

Figure 4-3 illustrates the scoring algorithm for Residential Upstream Lighting Free Ridership via In-Store Intercepts.

⁴³ ComEd has used this method since EPY2. Ameren began using it in EPY5.

Figure 4-3. Residential Upstream Lighting Free Ridership via In-Store Intercept

4.3.1.2 Data Collection

To estimate free ridership, the evaluation teams will conduct in-store intercept surveys with customers purchasing program-discounted lighting at participating retailers. Customers are asked questions that are used to estimate a Program Influence Score and a No-Program Score for each customer and efficient bulb type purchased.

Primary Program Influence Score Questions

1. Light bulb purchasing plans for current shopping trip (Yes/No)
2. If planning to purchase bulbs:
 - a. Bulb type (CFL, LED, Incandescent, Halogen)
 - b. Program administrator-incentivized bulbs (Yes/No)
3. Influence of various program factors:
 - a. Program incentive
 - b. In-store information (printed materials or information from Program Administrator representatives or retail personnel)
 - c. Positioning of discounted bulbs within the store

Primary No-Program Score Questions

1. Stated preference of light bulb purchases had the Program Administrator incentive not been available (purchase all, some, or none of efficient bulbs)
2. Quantity of light bulbs purchased absent the incentive

4.3.1.3 Scoring Algorithms

Using the data collected from program participants during the in-store intercept surveys, Program Influence and No-Program Scores are calculated for each survey respondent and then combined to estimate a respondent-specific Free Ridership Score.

4.3.1.3.1 Calculation of the Program Influence Score

Survey respondents purchasing one or more program-discounted bulbs are assigned a Preliminary Program Influence Score based on the maximum program influence level (on a 0 to 10 scale) they assigned to one or more program factors (e.g., monetary incentive/informational materials [printed or from store personnel]/product positioning). The influence level assigned to the monetary incentive should be increased for survey respondents (using a linear decreasing function)⁴⁴ who indicated that, absent the incentive, they would not have purchased any of the program bulbs they were purchasing that day.

After the Preliminary Program Influence Score is assigned, a secondary algorithm is run that adjusts the preliminary program influence based on survey data regarding the customers purchasing plans when they entered the store. Survey respondents who indicated they planned to purchase high-efficiency bulbs prior to entering the store and who had not come to the store specifically to buy Program Administrator-incentivized program bulbs, should have their Program Influence Score cut in half. This adjustment makes the final Program Influence Score reflective of their stated planned intention to purchase efficient bulbs in the program's absence.

4.3.1.3.2 Calculation of the No-Program Score

The No-Program Score is based on whether a respondent states they would have purchased all, some, or none of the program-discounted bulbs in the absence of Program Administrator incentives. Respondents reporting they would have purchased all of the efficient bulbs without the incentive should be considered free riders and receive a No-Program Score of zero. Those reporting they would have purchased none of the efficient bulbs without the incentives should be classified as non-free riders and receive a No-Program Score of 10, the maximum. Respondents reporting they would have purchased some of the efficient bulbs without the incentive should be assigned a No-Program Score between 0 and 10, reflective of the percentage of efficient bulbs they would not have purchased absent the program.

Respondents reporting they would have purchased all of the program-discounted bulbs in the program's absence, but in-store materials provided by the Program Administrator had a moderate to high influence on their decision, should have their No-Program Scores adjusted to equal the level of influence they attributed to these program-sponsored informational materials.

4.3.1.4 Calculation of Free Ridership

The Free Ridership rate is calculated as follows:

$$\text{Free Ridership} = 1 - (\text{Program Influence Score} + \text{No-Program Score})/20$$

Using the calculated Program Influence and No-Program Scores, Free Ridership is calculated as one minus the sum of the two scores (Program Influence Score plus No-Program score), divided by 20. Dividing the sum of scores by 20 results in a ratio (between 0 and 1) that is representative of the average of the two zero to 10 scores. Subtracting this ratio from one reverses the score, thus representing the free ridership level. If either the No-Program or Program Influence Scores are missing, Free Ridership can be calculated using the single available score divided by 10. Evaluators may also reference available data to perform documented modifications to individual free ridership estimates resulting from the application of this free ridership assessment methodology.

⁴⁴ The function, adjusted monetary score = (monetary score + 10)/2, increases the monetary score using a decreasing linear function. This function results in an increase in the monetary influence score of between 0 and 5 points depending on their original monetary score (i.e., an original score of 0 would become a 5, a 5 would become a 7.5, and a 10 would remain a 10). In past Illinois evaluations, this adjustment has typically changed less than 10% of all monetary scores.

4.3.2 Participant Spillover

For this program, participant spillover results from purchases of non-discounted efficient bulbs by program bulb purchasers who are influenced by their participation in the residential lighting program to purchase additional non-discounted efficient bulbs.

4.3.2.1 Data Collection

Data collected during in-store intercept surveys with customers purchasing program bulbs should be used to estimate participant spillover. During these surveys, customers purchasing program-discounted and non-discounted efficient bulbs should be asked questions to determine whether the residential lighting program influenced their purchases of non-discounted efficient bulbs.

4.3.2.2 Scoring Algorithm

To estimate participant spillover, the number of program-influenced, non-discounted efficient bulbs purchased by program participants is divided by the total number of program bulbs purchased by these program participants. This results in the Participant Spillover Rate.

4.3.3 Nonparticipant Spillover

Nonparticipant spillover results from purchases of non-discounted efficient bulbs by customers who are not purchasing program-discounted bulbs, but report that the residential lighting program influenced their decision to purchase non-discounted efficient bulbs.

4.3.3.1 Data Collection

Data collected during in-store intercept surveys with customers purchasing efficient bulbs not discounted by the program should be used to estimate nonparticipant spillover. During these surveys, customers purchasing non-discounted efficient bulbs should be asked questions to determine whether the residential lighting program influenced their purchases of non-discounted efficient bulbs.

4.3.3.2 Scoring Algorithm

To estimate nonparticipant spillover, one must first calculate the number of program-influenced, non-discounted efficient bulbs purchased by the population of program nonparticipants surveyed. This yields a survey nonparticipant spillover rate. This rate is then extrapolated to the estimated population of nonparticipating customers to determine the estimated total quantity of non-program efficient bulbs being purchased within the utility service territory. Dividing this result by the total number of program bulbs results in the Nonparticipant Spillover Rate.

4.3.3.3 Method Advantages and Disadvantages

The in-store intercept method described above has certain advantages and disadvantages.

Advantages: This approach catches customers at their point of purchase, before they leave the store and can no longer be contacted directly. Given the interview's timing, customers can more easily recall price factors leading to their purchase choices. Also, as customers are intercepted at the store rather than surveyed by telephone, a higher cooperation rate results.

Disadvantages: Customers may not fully connect the impact that in-store education, product placement, and advertising have on their decision making. While many consumers believe they are not influenced by advertising, retailers know advertising and product placement work. Further, store intercepts typically must be coordinated with education events, and many retailers do not allow interviews to take place in their stores. Consequently, results are not based on random samples of customers purchasing program-discounted lighting throughout the year and across all participating retailers, which could bias the results.

4.4 Prescriptive Rebate (With No Audit) Protocol

Prescriptive Rebate programs typically offer predetermined rebates to residential customers for purchasing measures such as high-efficiency furnaces, clothes washers, brushless/electronically commutated motors (ECMs), boilers, boiler reset controls, water heaters, air-source heat pumps (ASHPs), ground-source heat pumps (GSHPs), central air conditioners (CACs), programmable thermostats, smart thermostats, insulation, air sealing, duct sealing, and desktop power management software. The program may require installation by a registered program ally, but it does not require a home audit (although purchases may be made in response to an audit).

These programs encourage consumers to undertake the following:

- Purchase higher-efficiency equipment than they otherwise would have, had they shopped for such equipment at the same time (replace on burnout); and
- Replace operating but inefficient equipment with higher-efficiency equipment (early replacement).

The basic method for estimating free ridership and participant spillover (See Section 4.1.2) for these programs uses a participant self-report, based on a standard battery of questions. An enhanced method may utilize trade ally surveys to provide another quantitative assessment, which may be triangulated with the basic method approach. As discussed further in Section 4.4.2, trade ally surveys may also be used to assess nonparticipant spillover.

4.4.1 Basic Method

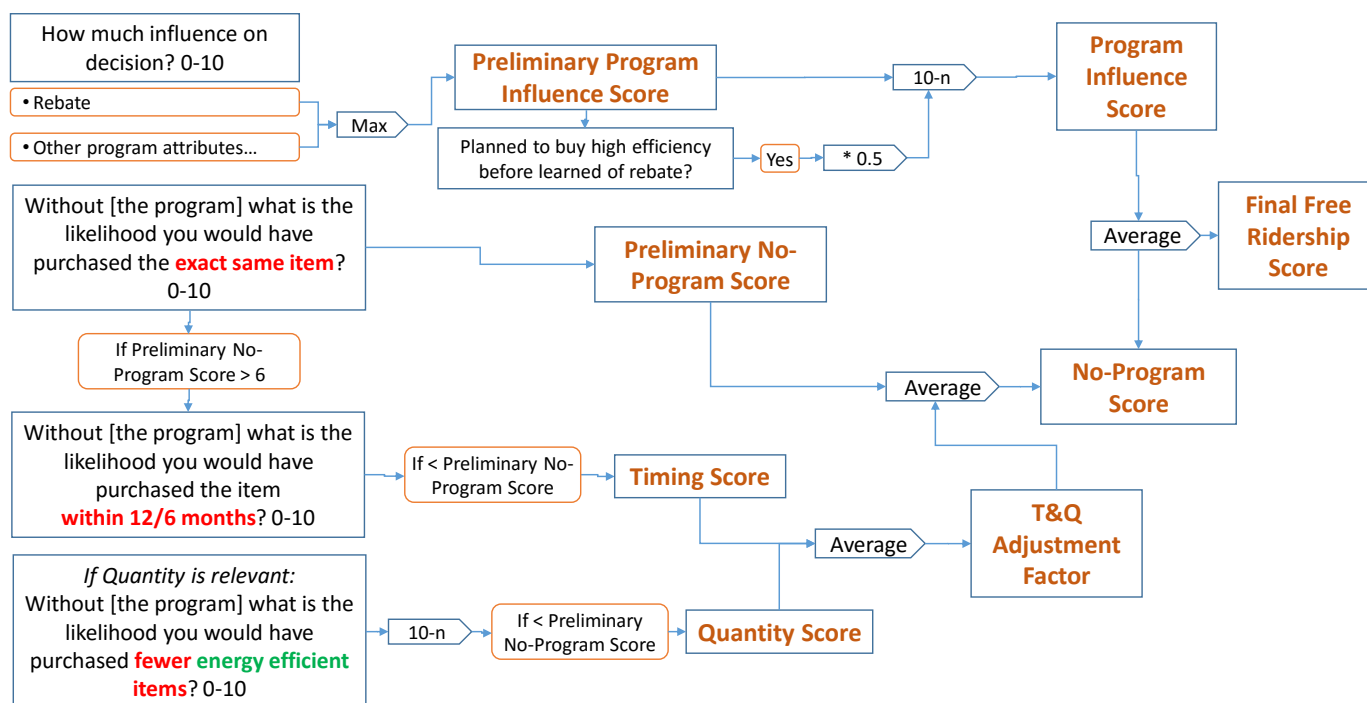
4.4.1.1 Free Ridership

The free ridership assessment battery is brief to avoid applying an undue survey burden, yet it seeks to reduce self-report biases by including two main free ridership components:

- A Program Influence component, based on the participant's perception of the program's influence on the decision to carry out the energy-efficient project; and
- A No-Program component, based on the participant's intention to carry out the energy-efficient project without program funds.

When scored, each component assesses the likelihood of free ridership on a scale of 0 to 10, with the two scores averaged and for a combined total free ridership score. For replace on burnout, adjustments are possible for timing and quantity impacts, assessed through additional questions, while for early replacement, timing is a major driver of the assessment. As different and opposing biases potentially affect the two main components, the No-Program component typically indicates higher free ridership than the Program Influence component. Therefore, combining these decreases the biases.

If the program differentiates between replace on burnout and early replacement and pays different incentives, the free ridership protocol applicable to the type of replacement should be followed. Figure 4-4 illustrates the scoring algorithm for Replace On Burnout measures, while Figure 4-5 illustrates the scoring for Early Replacement. Should the program not differentiate between replacement types, the Replace On Burnout protocol should be followed.

Figure 4-4. Residential Prescriptive Rebate (With No Audit) Free Ridership (Replace on Burnout)

4.4.1.1.1 Calculation of the Program Influence Score

Program influence is assessed by asking respondents, on a scale from 0 (not at all important) to 10 (very important), how important they found various program elements were on their decision to undertake the project the way they did. The number of elements included will vary, depending on the program's design. Logic models, program theory, and staff interviews typically inform the list of elements. Programs typically use the following elements to influence customer decision making: information; incentives or rebates; interaction with program staff (i.e., technical assistance); interaction with program proxies, such as members of a trade ally network; building audits or assessments; and financing.

In addition to asking about specific program influences, surveys ask respondents whether they planned to purchase a high-efficiency version of the product before learning of the rebate program. The Preliminary Program Influence Score is adjusted by 0.5 for those answering the question "yes."⁴⁵ Evaluators should conduct a sensitivity analysis around the use of this adjustment and present it in the report.

The Preliminary Program Influence Score equals the maximum influence rating for any program element rather than, for example, the mean influence rating. This is based on the rationale that if any given program element had a great influence on the respondent's decision, then the program itself had a great influence, even if other elements had less influence.

⁴⁵ The Illinois NTG Working Group discussed using this question to check for consistencies rather than adjusting the score. The NTG working group agreed that it is preferable not to directly ask about conflicting language with residential customers and to utilize an open ended question instead to assess possible reasons for conflicting statements. It is the experience of the NTG working group members that residential customers tend to be more impatient with these types of questions and can typically respond easier to an open-ended question about their motivations.

An inverse relationship occurs between high program influence and free ridership: the greater the program influence, the lower the free ridership. The Program Influence (PI) Score = 10 - Preliminary Program Influence Score. Should a respondent also answer that they planned to purchase a high-efficiency version of the product before learning of the program, the Preliminary Program influence Score should be multiplied by 0.5 before subtracting from 10 to compute the Program Influence Score.

4.4.1.1.2 Calculation of the No-Program Score

The No-Program (NP) Score differs slightly, depending on whether it is a Replace On Burnout measure or early replacement. For Replace On Burnout, the Preliminary No-Program (NP_p) Score is assessed using a key question that asks the respondent to gauge their likelihood of purchasing the exact same item (e.g., make, model, efficiency) had the program not existed. A respondent stating the likelihood of purchasing the same exact item as a 5 on a scale of 0 to 10 is assigned a Preliminary No-Program Score of 5.

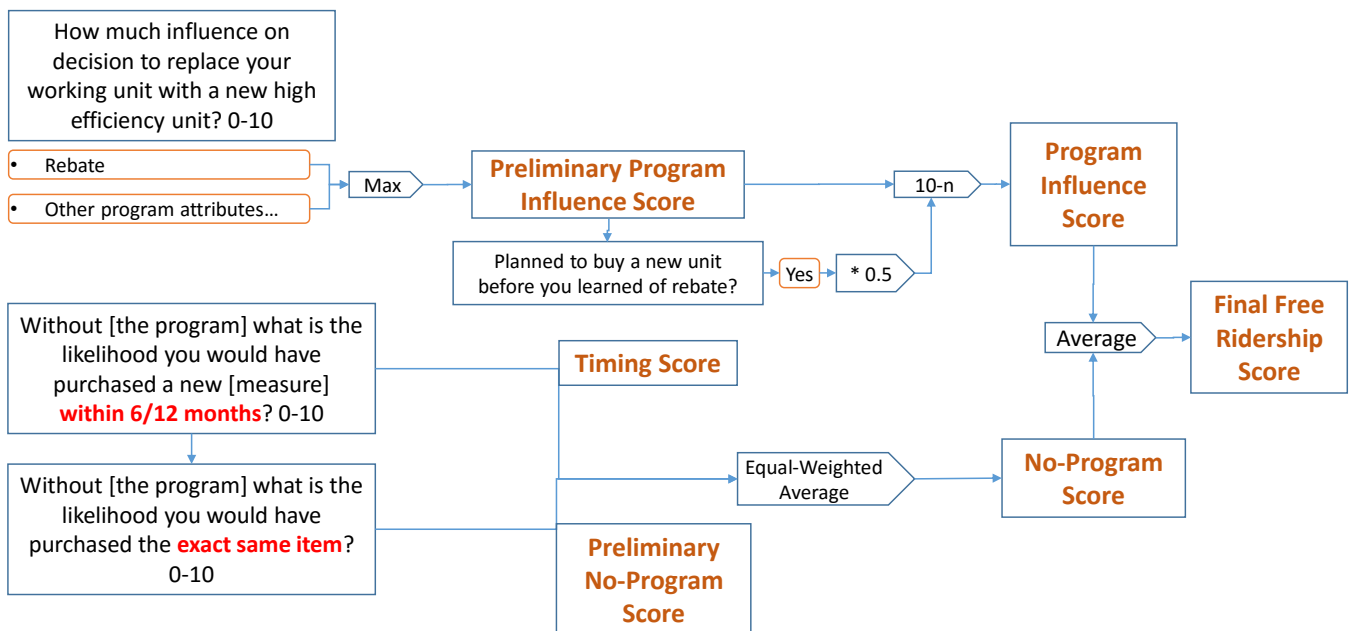
Two possible adjustments to the Preliminary No-Program Score address the possibility of the program accelerating respondents' decisions to purchase or increasing the quantity of efficient products purchased (i.e., partial free ridership) compared to actions taken in the program's absence. These adjustments are based on the responses to two additional questions.

If the Preliminary No-Program Score is greater than 6, the participant should be asked their likelihood of purchasing the same item within 12 or 6 months (12 months for a single or big ticket item and 6 months for less expensive items) for the Timing (T) Score.

If multiple quantities of an item are purchased, the respondent should be asked about the likelihood of purchasing fewer energy-efficient items. The response to this question is subtracted from 10 to compute the Quantity (Q) Score.

If either or both of these scores (timing or quantity) is less than the Preliminary No-Program Score, the T&Q Adjustment Factor (average of Timing and Quantity Scores) should be averaged in with the Preliminary No-Program Score to get the No-Program Score. The No-Program Score is averaged with the Program Influence Score to calculate the Final Free Ridership Score.

Figure 4-5. Residential Prescriptive Rebate (With No Audit) Free Ridership (Early Replacement)



For early replacement, the No-Program score is assessed by averaging the likelihood ratings from two questions:

- Likelihood of purchasing a new measure within 12 months (Timing Score); and

- Likelihood of purchasing the exact same item (Preliminary No-Program Score).

The two likelihood scores are weighted equally and result in the No-Program score, which is then averaged with the program influence score to calculate the final free ridership score.

In addition to reporting results based on the algorithms outlined above, evaluators should test two alternative methods of combining the scores and then should report the sensitivity of results to these changes. For Replace On Burnout measures, evaluators should test the following scenarios:

1. An average of Preliminary No-Program (NP_p), Timing (T), and Quantity (Q) Scores, which is then averaged with the Program Influence (PI) Score (if applicable):

$$FR = Mean\left(Mean(NP_p, T, Q)\right), PI$$

2. Average of all four scores, PI, NP_p , T, and Q:

$$FR = Mean(PI, NP_p, T, Q)$$

Similarly, for Early Replacement measures, evaluators should test the following scenario:

- Average of Preliminary No-Program (NP_p), Timing (T), and Program Influence (PI) Score:

$$FR = Mean(PI, NP_p, T)$$

Evaluated net savings results will be based on the algorithm outlined in this protocol, but sensitivity analysis results may be used to develop suggested algorithm revisions going forward.

4.4.1.1.3 Consistency Checks

To address the possibility of conflicting responses (i.e., low intention score and high influence score), the survey should include consistency checks that, at a minimum, ask participants an open-ended question to address the program's influence. For example:

- In your own words, please tell me the influence the program had on your decision to purchase the <insert measure name>.

In this case, the evaluation analyst will assess the response to this open ended question and its consistency with the other questions, and, if warranted based on clear additional information, they will adjust the score based on expert judgement. If an inconsistency exists and the open-ended response does not resolve the inconsistency, the respondent will be removed from the calculation. All instances of this occurring should be documented in the final report. Additional consistency checks, triggered and resolved within the survey with additional questions to participants, remain optional.

Missing responses to specific questions should be treated as "missing" for that particular question, but the observation or case will be retained in the analysis. Evaluation reports should note if this affects more than 5% of the responses.

4.4.2 Enhanced Method

4.4.2.1 Free Ridership

Free ridership results may be enhanced by including additional research efforts. A trade ally survey can be conducted to assess the percentage change in sales of high-efficiency equipment resulting from the program and the percentage of efficient equipment sales rebated through the program. Though these questions avoid directly asking for total sales before and after the program, the "with program" sales volume can be calculated by dividing program tracking database counts of rebated products by the percentage of efficient products rebated through the

program. The “without program” sales volume would then equal the “with program” sales volume, adjusted by the reported percentage change in equipment sales resulting from the program. Evaluators should ensure that trade allies receive sufficient time to collect specific data and not rely on “guesses” to respond. These results may be triangulated with participant survey results.

4.4.2.2 Triangulation

When multiple methods are used, evaluators may triangulate results by rating the analysis methodology and data collected using responses (rated on a scale of 0 to 10) to the following three questions:

1. All things being equal, on a scale of 0 to 10, with 0 being not at all likely and 10 being extremely likely, how likely is the approach to provide a more accurate estimate of free ridership?
2. Similarly, how valid is the data collected and the analysis performed (i.e., consider missing data, whether data collected was based on recollection or record keeping, is the analysis technique able to properly utilize the data collected)?
3. How representative is the sample (accounting for confidence and precision, and non-response or any sample frame bias)?

The weight for each method is the average score for that method divided by the sum of the scores for all methods.

Table 4-5 illustrates example scoring for two different methods, illustrating the calculated weights.

Table 4-5. Example Triangulation Weighting Approach

NTG Triangulation Data and Analysis	Method 1	Method 2
1. How likely is this approach to provide an accurate view of free ridership?	6	8
2. How valid is the data collected/analysis?	3	5
3. How representative is the sample?	8	10
Average Score	5.7	9
Weight	39%	61%

4.5 Single-Family Home Energy Audit Protocol

Single-Family Home Energy Audit programs (or energy assessment programs) seek to secure energy savings for residential customers by providing audits, direct-install measures, and incentives for additional energy efficiency opportunities. The participation process generally begins with an energy audit, performed by a program-affiliated companies or individuals; this involves an auditor assessing the customer’s home to identify energy-saving opportunities. At that time, the auditor may install free instant-savings measures, such as CFLs, low-flow showerheads, and faucet aerators. Auditors also may educate customers about incentives available through the audit program (e.g., air sealing, insulation) or other Program Administrator-sponsored energy efficiency programs.

For these programs, free ridership and participant spillover (See Section 4.1.2) estimates rely on participant self-reports, gathered through surveys.

4.5.1 Basic Method

Given the multiple components of some audit programs, net impacts should be estimated using survey batteries tailored to a customer’s experience (e.g., receipt of free direct-install measures and discounted or rebated measures). The following sections outline the approach for two program components, one dealing with the direct installation of free low-cost measures and a second dealing with envelope measures, such as air sealing and insulation.

4.5.1.1 No-Cost, Direct Install Measures

For free measures directly installed by program staff due to the audit, free ridership calculations should include the following components: No-Program, Timing, and Quantity.

This approach provides several important benefits, such as deriving a partial free ridership score based on the likelihood that the participant would take similar actions in the absence of the audit. For example, partial scores can be assigned to customers who planned to install the measure, but the program influenced that decision, particularly in terms of timing (e.g., the program might have accelerated the installation) or quantity (e.g., the program might have led to installation of additional program-qualified measures).

Outlines of components and their associated survey questions follow:

- **No-Program Score (NP).** This score reflects the likelihood that customers would have installed specified measures, had the program not existed. For free measures, this is based on a question asking respondents to rate the likelihood that they would have installed the exact same measures had they not received them for free through the audit (on a 0 to 10 scale, where 0 is not at all likely and 10 is extremely likely). A higher likelihood value means a higher level of free ridership (i.e., a lower attribution level for the program).
- **Timing and Quantity Adjustment Factor (T&Q_{ADJ}).** This factor adjusts the No-Program Score downwards for an earlier installation of equipment and for the installation of greater equipment quantities than would have taken place in the program's absence. It is based on two questions targeting respondents who report they would likely have installed the same equipment without the program (a score greater than 6, on a scale from 0-10, on the NP question). This will only be asked of measures where the quantity is a relevant concept (e.g., lighting).

The first question to compute the Timing (T) Score accounts for earlier installation of measures due to the program by asking respondents about their likelihood (0-10 scale) to have installed the measure within 6 or 12 months, had they not received it through the program (12 months for a single or big ticket item and 6 months for less expensive items).

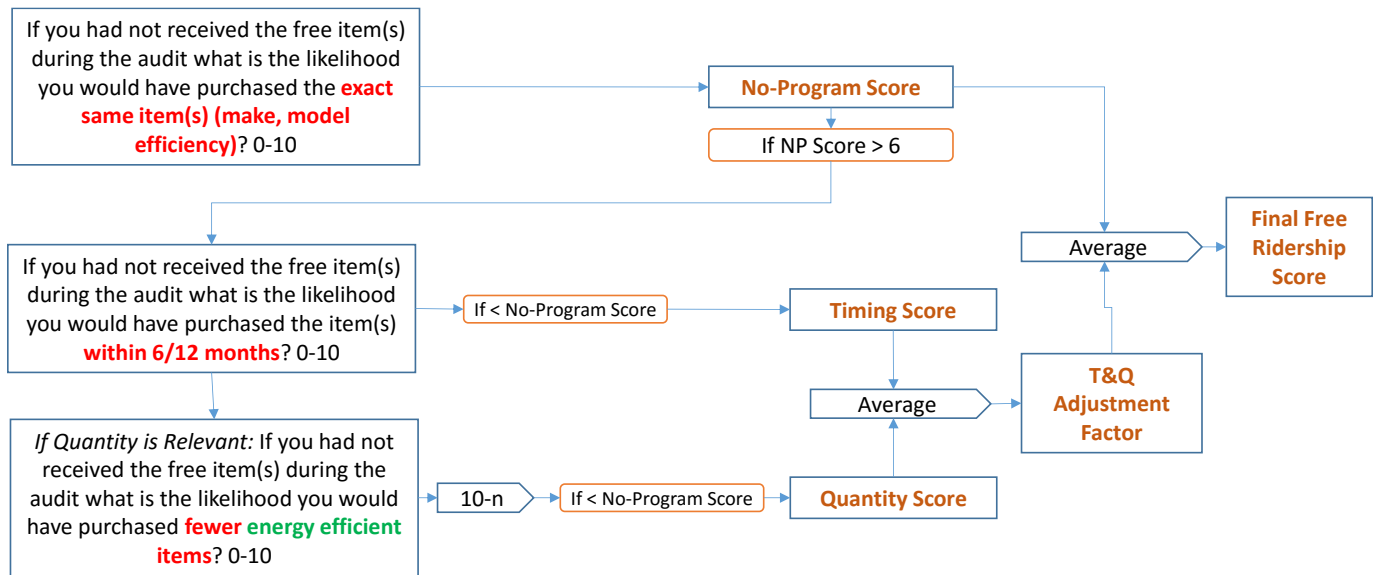
The second question to compute the Quantity (Q) Score asks respondents about the likelihood that they would have installed fewer measures or performed less weatherization without the program. The response to this question is subtracted from 10 to compute the Quantity Score, as a lower score means a greater likelihood respondents would have installed the same or a greater number of measures. Averaging the Timing and Quantity scores provides the adjustment factor, as shown in the following equation:

$$T\&Q_{ADJ} = \text{Mean}(T, Q)$$

Given the low cost of the measures provided through the direct-install component of most audit programs and the number of measures received per participant, efforts have been made to streamline the free ridership battery to reduce the respondent's burden. As such, the overall Final Free Ridership Score per measure can be calculated by averaging the NP Score and the Timing and Quantity Adjustment Factor, as shown in the following equation:

$$\text{Free Ridership (FR)} = \text{Mean}(NP, T\&Q_{ADJ})$$

Figure 4-6 illustrates the algorithm for no cost measures.

Figure 4-6. Single-Family Home Energy Audit Free Ridership—No Cost Measures

4.5.1.2 Rebated/Discounted Measures

Estimating NTG for rebated measures (typically for building shells) requires a more rigorous process than estimating NTG for free direct-install measures. In particular, the approach integrates an assessment of various program components that may have influenced the participant's decision to install the measures. For discounted envelope measures, the basic free ridership factor consists of the following two components:

- **Program Influence (PI).** Evaluators assess program influence by asking respondents how important various program elements were on their decisions to undertake a project as they did (on a scale from 0 to 10, where 0 is not at all important and 10 is very important). The list of program elements draws upon the program's design and theory as well as input from program staff. In addition to asking about specific program influences, surveys ask respondents whether they installed the equipment or performed a particular upgrade before learning about the program. The Preliminary Program Influence Score is adjusted by 0.5 for those answering "yes" to this question. Evaluators should conduct a sensitivity analysis around the use of this adjustment and present it in the report.

The Preliminary Program Influence Score equals the maximum influence rating for any program element rather than, for example, the mean influence rating. This is based on the rationale that if any given program element had a great influence on the respondent's decision, then the program itself had a great influence, even if other elements had less influence.

An inverse relationship occurs between high program influence and free ridership: the greater the program influence, the lower the free ridership. The Program Influence (PI) Score = 10 - Preliminary Program Influence Score. Should a respondent also answer that they planned to purchase a high-efficiency version of the product before learning of the program, the Preliminary Program Influence Score should be multiplied by 0.5 before subtracting from 10 to compute the Program Influence Score.

- **No-Program Score (NP).** The Preliminary No-Program (NP_p) Score reflects the likelihood that customers would have installed the specified measures in the program's absence. In general, this draws upon one question that asks respondents to rate the likelihood that they would have installed the exact same measures if not receiving them through the program (on a 0 to 10 scale, where 0 is not at all likely and 10 is extremely likely). A higher likelihood value means a higher free ridership level (i.e., a lower attribution level for the program).
 - **Timing and Quantity Adjustment Factor (T&Q_{Adj}).** This factor adjusts the Preliminary No-Program Score downwards for earlier installation of equipment and for installation of greater

equipment quantities than would have taken place in the program's absence. It draws upon two survey questions asked of respondents who indicate they likely would have installed the same equipment without the program (a score greater than 6, on a scale from 0-10, on the NP question).

The first question to compute the Timing (T) Score accounts for earlier measure installation due to the program, asking respondents about the likelihood (0-10 scale) that they would have installed the measure within 6 or 12 months (12 months for a single or big ticket item and 6 months for less expensive items) had they not received it through the program.

The second question asks respondents about their likelihood to have installed fewer measures or performed less weatherization in the program's absence. Responses to this question are subtracted from 10 to compute the Quantity (Q) Score, as a lower score means a greater likelihood that they would have installed the same or a greater number of measures. This will only be asked of measures where the quantity is a relevant concept.

Averaging the Timing and Quantity Scores produces the adjustment factor, as shown in the following equation:

$$T\&Q_{ADJ} = \text{Mean}(T, Q)$$

Averaging the Timing and Quantity Adjustment Factor with the Preliminary NP Score (NP_p) calculates the final No-Program (NP) Score:

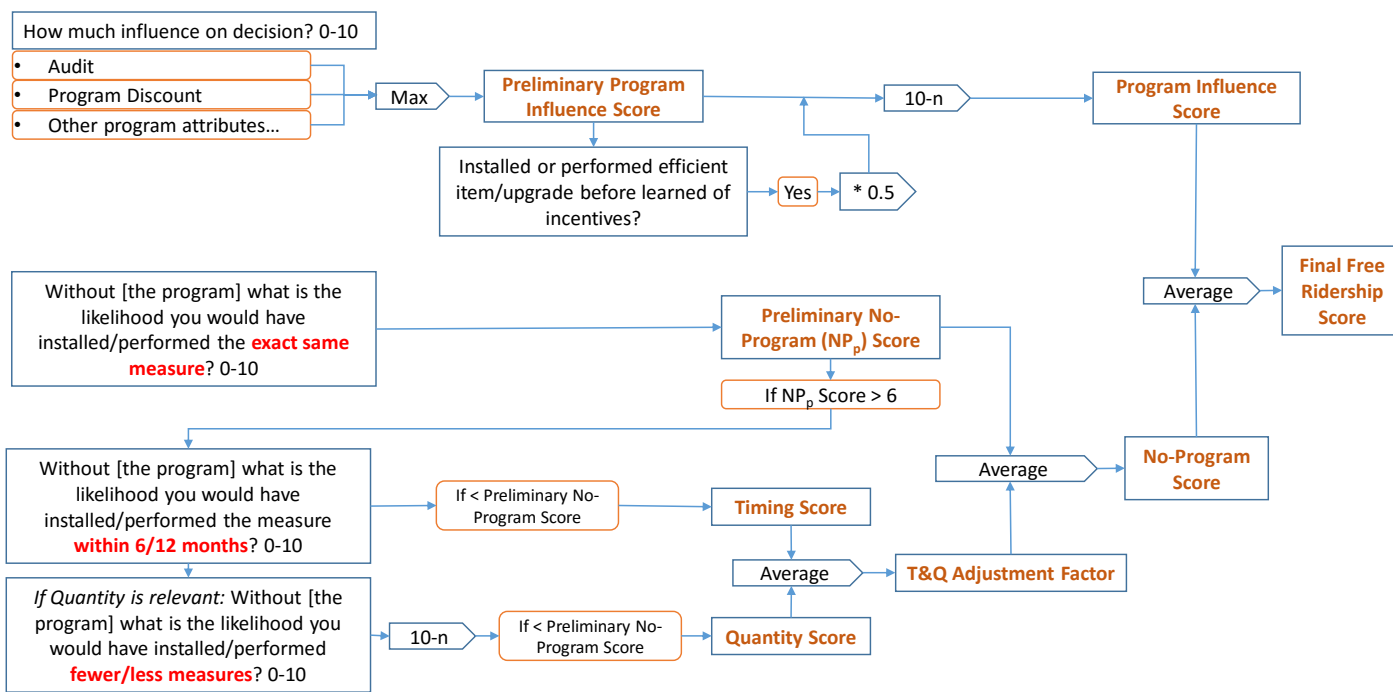
$$NP = \text{Mean}(NP_p, T\&Q_{ADJ})$$

Averaging the PI and NP Scores produces the overall free ridership score, as shown in the following formula:

$$\text{Free Ridership (FR)} = \text{Mean}(PI, NP)$$

Figure 4-7 illustrates the algorithm for discounted measures.

Figure 4-7. Single-Family Home Energy Audit Free Ridership—Discounted Measures



In addition to reporting results based on the algorithms outlined above, evaluators should test two alternative

methods of combining the program influence, no-program, timing, and quantity scores, and then should report the sensitivity of results to these changes. In particular, evaluators should test the following scenarios:

1. An average of NP_p, T, and Q scores, which is then averaged with the PI score (if applicable):

$$FR = \text{Mean}(\text{Mean}(NP_p, T, Q), PI),$$

2. Average of all four scores, PI, NP_p, T, and Q:

$$FR = \text{Mean}(PI, NP_p, T, Q)$$

Evaluated net savings results will be based on the algorithm outlined in this protocol, but sensitivity analysis results may be used to develop suggested algorithm revisions going forward.

4.5.1.3 Consistency Checks

To address the possibility of conflicting responses (e.g., the high likelihood to install the same measure in the program's absence and the high importance to program factors), the survey should include consistency checks that, at a minimum, ask participants an open-ended question to address a program's influence, such as the following:

- In your own words, please tell me the influence the program had on your decision to purchase the <insert measure name>.

For low or no-cost, direct-install measures, surveys should include two questions to assess a program's influence on the respondent. The first should be asked at the beginning of the NTG battery, and the second should be asked at its conclusion. Questions include the following:

- Prior to the audit, had you purchased any <measures>? Y/N
- IF YES AND LIKELIHOOD TO INSTALL WITHOUT THE PROGRAM IS <7: Given that you had purchased <measures> before receiving the audit, why didn't you purchase additional <measures> on your own without the program? [OPEN END]
- IF NO AND LIKELIHOOD TO INSTALL WITHOUT THE PROGRAM IS >6: Given that you have not purchased <measures> before, why were you likely to purchase <measures> on your own without the program? [OPEN END]

In both cases, the evaluation analyst will assess responses to open ended questions and their consistency with the other questions; if warranted, based on clear additional information, the evaluator will adjust the original question score will require adjustment. If inconsistency occurs and the open-ended response does not resolve it, the original question response will be removed from the calculation. Final reports should document all instances of such adjustments. Optionally, additional participant questions can be included to trigger and resolve additional consistency checks.

Missing responses to specific questions (e.g., don't know or refused) should be treated as "missing" for those particular questions, but the analysis retains the observation or case. The evaluation reports should note if this affects more than 5% of responses.

4.6 Multifamily Protocol

Multifamily energy efficiency programs typically offer direct installation of low-cost, energy-efficient measures in multifamily dwelling units, in addition to rebates for common area lighting retrofits, air sealing, insulation, and improvements to HVAC systems and controls. These programs have various target audiences from owners, managers, or developers of market rate multifamily housing to those operating lower income or assisted living housing. Across these groups, properties must generally have a minimum of between three and five units to

qualify for the programs.

Most multifamily program savings are typically achieved by encouraging customers to install higher-efficiency equipment than they would have installed on their own. However, programs may also encourage early replacement of still functioning equipment that is less efficient, thus impacting the timing of the installation, so that savings is realized earlier. The incentive may also make it more affordable for customers to install a greater number of high-efficiency measures.

The basic method for estimation of free ridership and participant spillover (See Section 4.1.2) for these types of programs is based on participant self-report gathered through surveys. For common area and building shell components of the program, participants are property managers and owners responsible for building maintenance and renovation. However, depending on the program design for the in-unit component of the program and specifically the installation of efficient lighting, the decision to participate in the program (i.e., install program measures) may arise from either property managers/owners or tenants or, potentially, both. This distinction is due to the fact that in some market-rate apartments, the tenant is responsible for decisions related to the installation of program measures, including light bulbs, while this is not common practice in income-qualified or assisted-living settings. For other in-unit measures, such as faucet aerators and low-flow showerheads, evaluators interview property managers/owners regarding program influence, as these measures are typically directly installed by program staff, and there is a limited likelihood of tenants making changes to these features.

To date, most programs have included CFLs as one of their measures; so the text in this section refers to CFLs. The protocol can also be applied when the program installs LEDs.

4.6.1 Basic Method

Free ridership is calculated as the average of two distinct components: a Program Influence Score and a No-Program Score, defined as follows:

- Program Influence component: based on the participant's perception of the program's influence on their decision to carry out the energy-efficient project; and
- No-Program component: based on the participant's intention to carry out the energy-efficient project without the program.

When scored, each component assesses free ridership likelihood on a scale of 0 to 10. Averaging and dividing the components by 10 produces a combined, total free ridership score, ranging from 0 to 100%. The two main components (i.e., No-Program and Program Influence) potentially are subject to different and opposing biases. Consequently, the No-Program component typically indicates higher free ridership than the Program Influence component. Combining these decreases the biases.

Descriptions of the algorithm components and their associated survey questions follow:

- **Program Influence (PI).** Evaluators assess program influence by asking respondents how important various program elements were in regard to their decisions to undertake the project as conducted (on a scale from 0 to 10, where 0 is not at all important and 10 is very important). The list is based on the program's design and theory as well as input from program staff. In addition to asking about specific program influences, surveys ask respondents whether they planned to install equipment or perform a particular upgrade before learning of the program. The Preliminary Program Influence Score is adjusted by 0.5 for those answering "yes" to that question. Evaluators should conduct a sensitivity analysis around the 0.5 adjustment factor and present it with the evaluation report.

The Preliminary Program Influence Score equals the maximum influence rating for any program element rather than, for example, the mean influence rating. This is based on the rationale that if any given program element had a great influence on the respondent's decision, then the program itself had a great influence, even if other elements had less influence.

An inverse relationship occurs between high program influence and free ridership: the greater the program influence, the lower the free ridership. The Program Influence (PI) Score = 10 - Preliminary Program Influence Score. Should a respondent also answer that they planned to purchase a high-efficiency version of the product before learning of the program, the Preliminary Program Influence Score should be multiplied by 0.5 before subtracting from 10 to compute the Program Influence Score.

- **No-Program Score (NP).** The Preliminary No-Program (NP_p) Score reflects the likelihood that the customer would have installed the specified measures in the program's absence. Generally, this is based on a question asking respondents to rate the likelihood that they would have installed the exact same measure without receiving them through the program (on a 0 to 10 scale, where 0 is not at all likely and 10 is extremely likely). A higher likelihood value means a higher free ridership level (i.e., a lower level of attribution to the program). Survey designers are encouraged to include questions in the property owner/manager survey to determine whether they routinely replace all bulbs (in specific areas) on a schedule to reduce maintenance costs. Such a pattern might help place answers in context.

For CFLs, the question asked of consumers differs slightly, given the specifics of program delivery. For example, program staff visit common areas, and, in some cases, tenant units, and replace working incandescent bulbs with CFLs. Thus, the CFL NP_p Score is based on a slightly different question: "If you had not received free CFLs through the program, how likely is it that you would have removed ANY of the working incandescent light bulbs in the common areas of your property and replaced them with CFLs? Please use a scale that ranges from 0 to 10, where 0 is "not at all likely" and 10 is "very likely."

- **Timing and Quantity Adjustment Factor (T&Q_{ADJ}).** This factor adjusts the Preliminary No-Program Score downwards for earlier installation of equipment and for installation of greater equipment quantities than those taking place without the program. This is based on two questions asked of respondents who say they would likely have installed the same equipment without the program (a score greater than 6, on a scale from 0-10, for the NP_p question). This will only be asked of measures where the quantity is a relevant concept (e.g., lighting).

The first question to compute the Timing (T) Score accounts for earlier installation of a measure due to the program by asking respondents about the likelihood (0-10 scale) that they would have installed the measure within 12 months had they not received the measure through the program.

The second question to compute the Quantity (Q) Score asks respondents the likelihood that they would have installed fewer measures or performed less weatherization in the program's absence. The response to this question is subtracted from 10 to compute the Quantity Score, as a lower score means a greater likelihood that they would have installed the same or a greater number of measures.

Averaging the Timing and Quantity Scores provides the adjustment factor, as shown in the following equation:

$$T\&Q_{ADJ} = \text{Mean}(T, Q)$$

The adjustment factor is then averaged with the Preliminary NP (NP_p) Score to calculate the No-Program (NP) Score:

$$NP = \text{Mean}(NP_p, T\&Q_{ADJ})$$

The overall free ridership score derives from the average of the PI and the NP Scores, as shown in the following formula:

$$\text{Free Ridership (FR)} = \text{Mean}(PI, NP)$$

Figure 4-8 and Figure 4-9 also illustrate the algorithms.

Figure 4-8. Multifamily Free Ridership for Property Managers—Non-CFL Measures

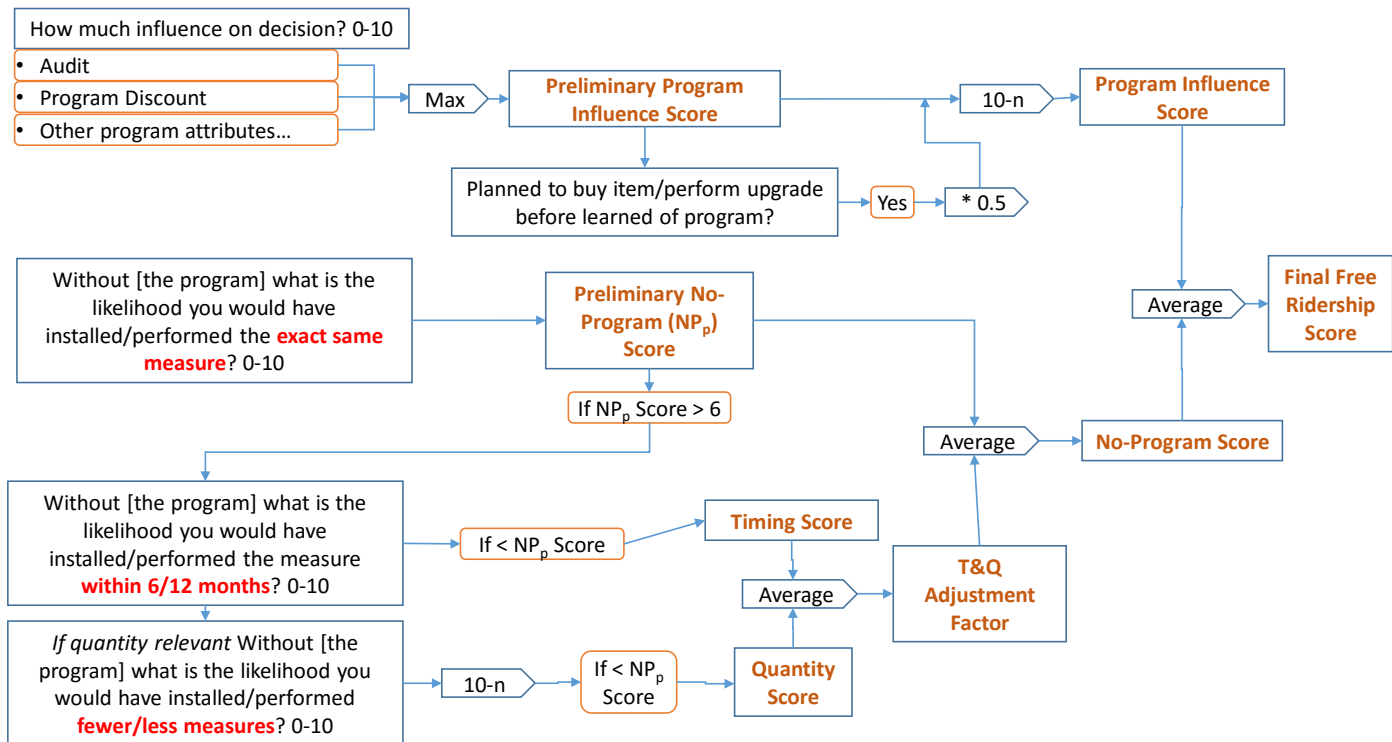
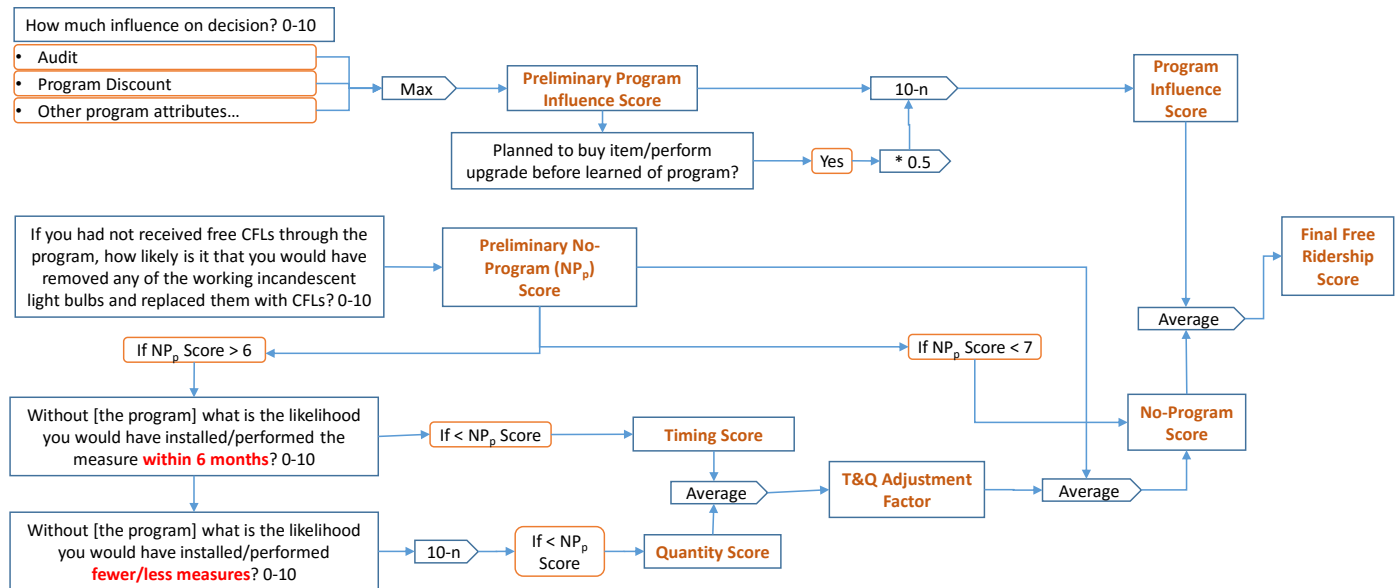


Figure 4-9. Multifamily Free Ridership—CFL Measures



In addition to reporting results based on the above algorithms, evaluators should test two alternative methods of combining the Program Influence, No-Program, Timing, and Quantity Scores, and reporting on the sensitivity of results to these changes. In particular, evaluators should test the following scenarios:

1. An average of the NP_p, T, and Q Scores, which is then averaged with the PI Score (if applicable):

$$FR = Mean(Mean(NP_p, T, Q), PI)$$

2. Average of all four scores—PI, NP_p, T, and Q:

$$FR = Mean(PI, NP_p, T, Q)$$

Though evaluated net savings results will be based on the algorithms outlined in this protocol, results from sensitivity analysis may be used to revise the algorithm going forward.

4.6.1.1 Consistency Checks

To address the possibility of conflicting responses (e.g., high likelihood to install the same measure without the program, high importance to program factors), the survey should include consistency checks that, at a minimum, ask participants an open-ended question to address the program's influence. For example:

- In your own words, please tell me the influence the program had on your decision to purchase the <insert measure name>.

The evaluation analyst will assess the responses to the open ended questions and their consistency with the other survey questions, and, if warranted based on clear additional information, will adjust the original question score. If the open-ended response does not resolve the inconsistency, responses to the original question should be removed from the calculation. The survey may include additional consistency check triggers and resolutions through additional participant questions. The final report should document how often the consistency check rules were triggered, how often adjustments were made to scores, and how often inconsistencies could not be resolved.

Missing responses to specific questions (including don't know or refused) should be treated as missing for that particular question, but the analysis should retain that observation or case. Evaluation reports should note if this affects more than 5% of the responses.

4.6.1.2 Data Collection

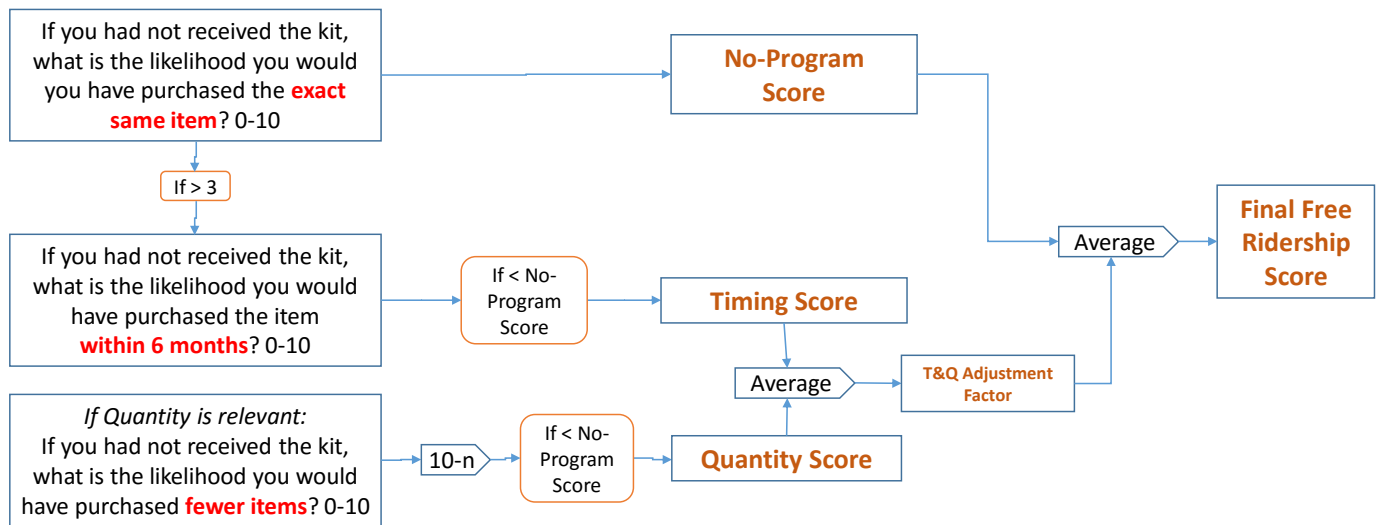
A participant survey should be used as the primary source of data collected for estimating free ridership in residential multifamily programs. As discussed, evaluators may field surveys with owners, property managers, or tenants, depending on a program's design and theory. Determining the appropriate audience from which to gather information for estimating free ridership depends on the program's design, and, ultimately, the party responsible for deciding to install specific program measures.

4.7 Energy Saving Kits and Elementary Education Protocol

Energy Saving Kits and Elementary Education Programs aim to secure energy savings through the distribution of kits containing various energy-saving measures, including (but not limited to): high-efficiency lighting (CFLs or LED lamps); bathroom and kitchen faucet aerators; and low-flow showerheads. Energy Saving Kits operate as an opt-in program; customers can request a kit by completing an Internet or phone application. Elementary Education Program participants do not request a kit as kits are distributed to all students in a classroom.

Free ridership and participant spillover (See Section 4.1.2) estimations for both programs rely upon participant self-report information gathered through surveys, despite the differences in distribution models. This methodology can be used for other energy-saving kit programs, including kits with alternative distribution methods (e.g., kits dropped off at a participant's home).

The following section contains a description of the basic NTG method used. Figure 4-10 illustrates the method.

Figure 4-10. Energy Saving Kits and Elementary Education Free Ridership

4.7.1 Basic Method

Free ridership calculations should include the following components: No-Program, Timing, and Quantity.

This approach provides several important benefits, such as the ability to derive a partial free ridership score based on the likelihood that similar actions would have taken place, even if the participant had not received a kit. For instance, partial scores can be assigned to customers with plans to install the measure, but the program at least influenced that decision, particularly in terms of timing (e.g., the program might have accelerated the installation) or quantity (e.g., the program might have led to the installation of additional measures).

An outline of components and their associated survey questions follows:

- No-Program Score (NP).** This score reflects what action the customer would have taken without the existence of the program. This derives from one question that asks respondents to rate the likelihood that they would have installed the exact same measure in the absence of receiving free measures in the kit (on a 0 to 10 scale, where 0 is not at all likely and 10 is extremely likely). A higher likelihood value means a higher free ridership level (i.e., a lower level of attribution to the program).
- Timing and Quantity Adjustment Factor (T&Q_{Adj}).** If the NP Score is greater than 3, the T&Q Adjustment Factor is computed. This factor adjusts the No-Program Score downwards for earlier installation of kit items and for installation of greater quantities than would have taken place in the program's absence. It draws upon two survey questions asked of respondents who indicate they likely would have installed the same equipment without the program (a score greater than 3, on a scale from 0-10, on the NP question).

The first question to compute the Timing (T) Score accounts for earlier installation of a measure due to the program's influence, and it derives from one question asking respondents about the likelihood (0-10 scale) that they would have purchased the measure within six months had they not received the measure in a kit.

The second question to compute the Quantity (Q) Score accounts for the installation of a greater number of measures than the customer would have installed independently. In particular, the question asks participants how likely they would have been to install fewer energy-efficient items had they not received them for free in the kit. The question applies only to measures with a quantity of greater than one. The response to this question is subtracted from 10 to compute the Quantity Score, as a lower score means a greater likelihood respondents would have installed the same or a greater number of measures.

Averaging the Timing and Quantity Scores provides the adjustment factor, as shown in the following equation:

$$T\&Q_{ADJ} = Mean(T, Q)$$

Given the low cost of measures provided in the energy-saving kits as well as the number of measures included in each kit, efforts have been made to streamline the free ridership battery to reduce the respondent's burden. Consequently, the resulting Timing and Quantity Scores are averaged to compute the $T\&Q_{ADJ}$ Factor, with that average then averaged with the No-Program Score to derive the overall Final Free Ridership Score per measure:

$$Free\ Ridership\ (FR) = Mean(NP, T\&Q_{ADJ})$$

In addition to reporting results based on the above algorithms, evaluators should test alternative methods for combining the No-Program, Timing, and Quantity Scores, and to report on the sensitivity of results to these changes. In particular, evaluators should test the following scenario:

- Average of NP, T, and Q Scores:

$$FR = Mean(NP, T, Q)$$

Evaluated net savings results will be based on the algorithm outlined in this protocol, but results from the sensitivity analysis may be used to suggest revisions to the algorithm going forward.

Missing responses to specific questions (e.g., don't know or refused) should be treated as "missing" for that particular question. Despite missing responses, the case will be retained in the analysis (pairwise deletion). The evaluation reports should present the percent missing for each of the three questions.

4.7.1.1 Data Collection

Evaluators should use a participant survey as the primary data collection source for estimating free ridership in Energy Saving Kits and Elementary Education Programs. As a general rule, a free ridership rate should be calculated for each separate kit component, and then be weighted by savings to determine the program-level results.

Free ridership questions should be asked near the beginning of a participant survey, before asking satisfaction questions. This should prevent participants from confusing free ridership questions with the satisfaction questions, which would result in an artificially high free ridership score, especially for participants highly satisfied with measures included in the kit.

4.8 Residential New Construction Protocol

Residential New Construction programs typically offer builder training, technical information, marketing materials, and incentives to builders for the construction of eligible homes. Eligible homes must meet specific standards, designed to achieve energy efficiency levels above local building codes. Programs may use different tiers of standards to meet correspondingly different incentives.

The basic method for estimating free ridership and participant spillover for these programs is based on builder participant self-reporting, gathered through surveys.

The following section describes the basic method used.

4.8.1 Basic Method

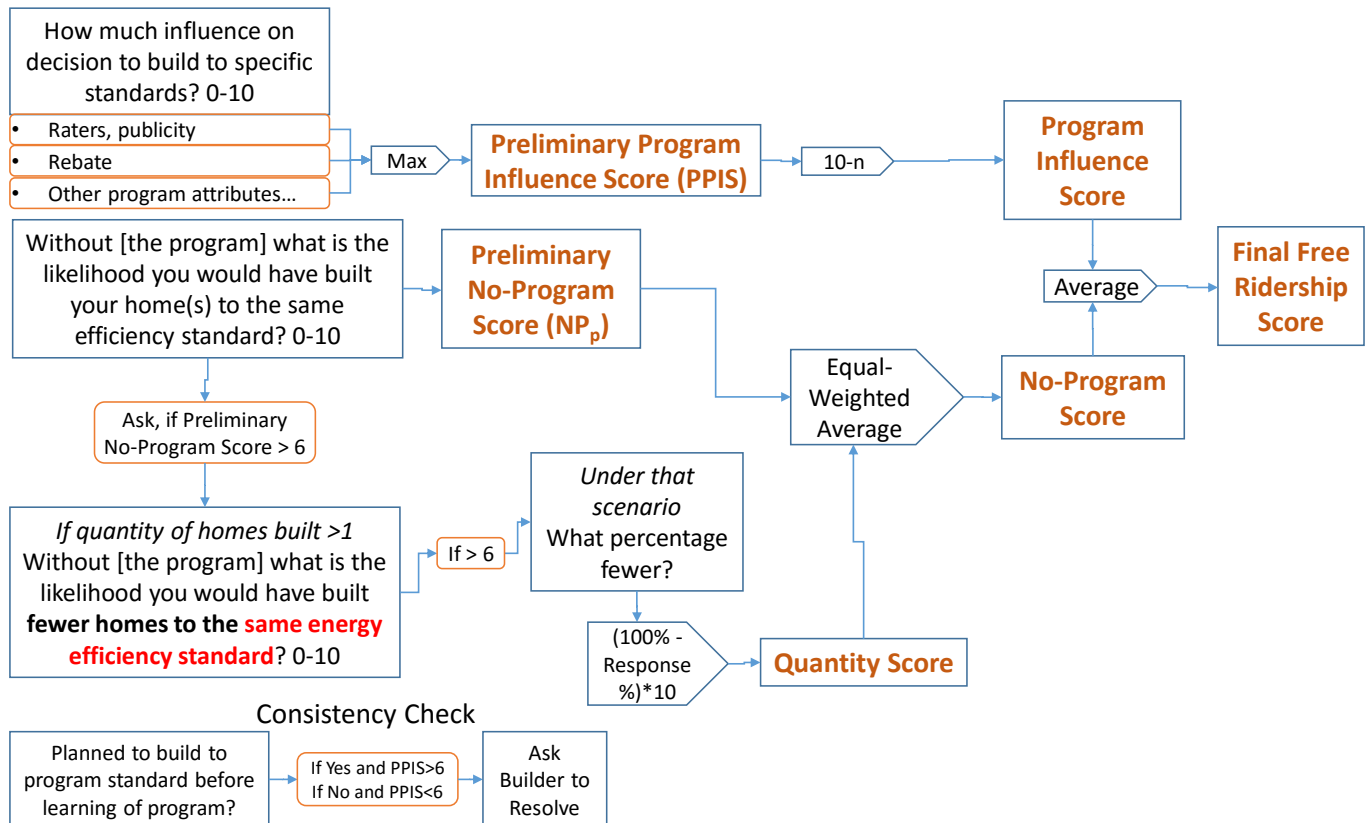
For this program, a free rider is a builder who would have constructed a home at the program's efficiency level in the program's absence. Given the multiple methods available to achieve desired home energy efficiency levels, survey questions consider the builder's likelihood of meeting the same energy efficiency standard, rather than whether or not the builder would have installed certain energy efficiency measures. Figure 4-11 (below) illustrates the method in more detail.

Evaluators assess Program Influence by asking respondents, on a scale from 0 (not at all important) to 10 (very important), how important they found various program elements in deciding to build to specific energy efficiency

standards. The number of elements included vary, depending on the program's design. Logic models, program theory, and staff interviews typically inform the list of program elements included. Programs typically use the following elements to influence builder decision making: marketing materials; incentives or rebates; contacts with HERS Raters; and technical assistance.

In addition to asking about specific program influences, surveys should ask builders whether they planned to build homes to the same standard before learning of the program.

Figure 4-11. Residential New Construction Free Ridership



4.8.1.1.1 Calculation of the Program Influence Score

The Program Influence Score (PI) equals 10 minus the maximum influence rating for any program element rather than, for example, the mean influence rating. This is based on the rationale that if any given program element had a great influence on the respondent's decision, the program itself had a great influence, even if other elements had less influence.

4.8.1.1.2 Calculation of the No-Program Score

Evaluators calculate the No-Program score using a set of questions that ask respondents to gauge their likelihood of building homes to the same standards and in the same quantities had the program not existed. Three separate responses are considered in calculating the No-Program Score:

- The likelihood, on a scale of 0 to 10, that the builder would have built their homes to the same efficiency standard (Preliminary No-Program Score (NPp))
- If that likelihood is greater than 6, the likelihood of fewer homes being built to the same efficiency standard.

- If that likelihood is greater than 6, the response to the question “for that scenario, what percentage of fewer homes would be built to the standard?” (Quantity Score = (100% - % answer) * 10, which will be a number between 0 and 10)

The resulting No-Program (NP) Score is calculated as follows:

$$NP = Mean(NP_p, Q)$$

The overall Free Ridership Score derives from the average of the PI and NP scores, as shown in the following formula:

$$FR = Mean(PI, NP)$$

4.8.1.2 Consistency Checks

To address the possibility of conflicting responses (e.g., the high likelihood to build to the same efficiency standards without the program, the high importance of program factors), the survey should include, at a minimum, consistency checks that ask participants an open-ended question to address the program’s influence. For example:

- In your own words, please tell me the influence the program had on your building practices.

If a high (>6) Preliminary Program Influence Score (PPIS) results, yet the builder planned to meet the same efficiency standard prior to learning of the program; or if the Preliminary Program Influence Score is lower (<7), and the builder did not plan to build to the standards prior to learning of the program, the survey should include a question to determine why this occurred, using wording that gets at the following inconsistencies:

- IF Preliminary Program Influence Score is >6 and Builder planned to meet the same efficiency standard prior to learning OF THE PROGRAM: Given that you had plans to meet the standard prior to learning about the program, why do you think the <program elements> were influential in your decision to meet the standard? [OPEN END]
- IF Preliminary Program Influence Score is <7 and Builder had no plans to meet the same efficiency standard prior to learning of the program: Given that you had no plans to meet the standard prior to learning about the program, why do you think the <program elements> were not more influential in your decision to meet the standard? [OPEN END]

The evaluation analyst will assess the responses to the open ended questions and their consistency with the other survey questions, and, if warranted based on clear additional information, will adjust the original question score. If the open-ended response does not resolve the inconsistency, responses to the original question should be removed from the calculation. The survey may include additional consistency check triggers and resolutions through additional participant questions. The final report should document how often the consistency check rules were triggered, how often adjustments were made to scores, and how often inconsistencies could not be resolved.

Missing responses to specific questions (including don’t know or refused) should be treated as missing for that particular question, but the analysis should retain that observation or case. Evaluation reports should note if this affects more than 5% of the responses.

4.8.2 Participant Spillover

Participant spillover occurs when, due to program participation, a builder increases the energy efficiency of homes built outside the program (but inside a utility's service territory) by adopting certain building practices used in participating homes. Participant spillover can be calculated based on participant builder survey questions that ask builders about homes built within the utility service territory but outside the program. Survey questions ask whether the builder increased the energy efficiency standards of non-program homes after participating in the program, and the number of homes they applied these increased standards to, within the utility's service territory. Depending on the program characteristics, spillover should be measured as changes in specific building practices or as installation of specific measures. The text below assumes the program has been targeted at modifying building practices.

Spillover may be recorded depending on responses to the following questions:

1. How important was your experience in the <PROGRAM ADMINISTRATOR'S> program in your decision to incorporate this building practice your other homes, using a scale of 0 to 10, where 0 is not at all important and 10 is extremely important?
2. If you had not participated in the <PROGRAM ADMINISTRATOR'S> program, how likely is it that you would still have incorporated this building practice using a 0 to 10, scale where 0 means you definitely WOULD NOT have implemented this practice and 10 means you definitely WOULD have implemented this practice?

Responses to the first question establish the Practice Attribution Score 1, and responses to the second question establish the Practice Attribution Score 2. Spillover may be program-attributable for building practices with self-report data meeting the following condition:

$$\text{Spillover Score} = (\text{Practice Attribution Score 1} + (10 - \text{Practice Attribution Score 2}))/2 > 7.0$$

For responses meeting these conditions, an evaluator determines that specific building practices referenced in the question are attributable to the program; otherwise, the evaluator determines that specific building practices referenced in the question are not attributable to the program. The attribution criteria represent a threshold approach, in which energy impacts associated with building practices program participants implement outside the program are either 100% program-attributable or 0% program-attributable.

For each building practice discussed, builders will be asked how they know the building practice is more efficient than other options. If the respondent can identify the building practice as ENERGY STAR or name an efficiency level that the evaluator confirms as above the minimum federal standard, or if they identify a technology that the evaluator can confirm is above the minimum federal standard, this counts towards participant spillover.

Finally, depending on the building practice cited by the builder, follow-up questions should ask customers to provide reasonable information to allow the evaluator to estimate the amount of savings using IL-TRM protocols, such as quantity of appliances or the location and amount of insulation.

To calculate the spillover energy and demand savings for these actions, further questions should be asked to assess the gross savings of the building practice, through the appropriate version of the IL-TRM, if available, and the number of homes to which it applied. To develop the Spillover Rate, the total energy and demand impacts from the sampled participants who implemented efficient building practices in other homes due to participation in the program is summed, and then this sum is divided by the total ex post sample energy and demand impacts:

$$\text{Participant Spillover Rate (PSO)} = \frac{\text{Sum of Energy or Demand from Additional EE Practices}}{\text{Sample Ex Post Gross Energy or Demand Impacts}}$$

The equation used to adjust the Core NTGR based on participant spillover is as follows:

$$\text{NTGR} = (1 - \text{FR} + \text{PSO})$$

4.8.2.1 Sample

The sample for a spillover survey should be a random sample of current and up to one year previous program participants. Regardless of the year of participation, spillover should be measured within the set of homes that were completed within 12 months of the survey date.

4.8.3 Builder Nonparticipant Spillover

In addition to participant free ridership and spillover, new construction programs may create NPSO through builders exposed to the program but not actually participating. Rather, they implement some or all of the efficiency measures incorporated through the program in order to compete with builders that are participating.⁴⁶ NPSO caused by builders can be determined by surveying two groups of builders:

- “Drop out” builders, who participated in the program previously but have not participated in the past 12 months.
- True nonparticipating builders that report they were aware of the program or that other builders were taking steps to improve new home efficiency, but had never participated.

Surveys ask nonparticipating builders if their knowledge of other builders’ increased focus on energy efficiency influenced their building practices and in what manner, to quantify the program’s impact on nonparticipating homes. The survey questions will first identify specific building practices that go beyond the implemented energy code for the specific jurisdiction in which the builder is active. Table 4-6 lists the latest building energy code in place for most areas of Illinois. Evaluators should make efforts to ensure the building code under enforcement for each jurisdiction is used as the baseline when evaluating spillover savings.

Table 4-6. IECC 2012 Building Energy Code

Component	IECC 2012
Thermostat	Heating 72F Cooling 75F Programmable Thermostat
Ceiling	U-0.026
Walls	U-0.057
Floors	U-0.033
Slab	R-10, 2ft
Windows	U-0.32
Infiltration	5ACH50
Duct Leakage	4CFM/100CFA
Duct Insulation	R-8 Attic Supply, R-6 Otherwise
Heat Pump	7.7 HSPF
Furnace	80 AFUE

⁴⁶ NPSO also can arise from nonparticipating customers as a direct result of general energy efficiency education and promotion efforts. A separate protocol addresses such NPSO. Care should be taken to ensure the different approaches do not double-count NPSO.

Component	IECC 2012
Boiler	82 AFUE
AC	13 SEER
Lighting	75% CFL
Appliances	RESNET Default
Gas Water Heat	0.59 EF
Electric Water Heat	0.91 EF

For each component that is more efficient than code, the following additional questions are asked:

1. How many homes did you sell in <period> that incorporated this upgrade?
2. Of these homes, how many would have incorporated this upgrade, had the <program> not existed?

Evaluators should ensure that nonparticipant builders receive sufficient time to collect specific data and not rely on “guesses” to respond. Responses should also clarify whether sales counts are specific to the utility service territory in question.

The following steps calculate the program’s nonparticipant builder spillover percentage:

1. Compute the difference between the total reported number of efficiency upgrades sold and the total that would have been sold in the program’s absence to obtain the total number of upgrades by type of upgrade for that builder.
2. Multiply the total net number of upgrades of each type sold by each surveyed builder by the average gross unit savings for each upgrade type.
3. Sum the result for each builder from the previous step, and weight the results by the ratio of the population of non-active builders to the sample to compute the total spillover energy over the program period.
4. Divide the spillover energy savings by program gross savings.

Should a general population survey be implemented for nonparticipant spillover, care should be taken to ensure spillover is not double-counted.

5 Cross-Sector Protocols

The following sections include protocols that may be applicable to programs in the residential as well as in the commercial, industrial, and public sectors. Table 3-1 Commercial, Industrial, and Public Sector Programs and Table 4-1 Residential and Low Income Programs present information regarding the applicability of these protocols to specific programs.

5.1 Behavioral Protocol

5.1.1 Randomized Controlled Trials

The SEE Action Network's recent monograph on evaluating residential behavioral energy efficiency programs⁴⁷ indicates most of these programs are designed as randomized controlled trials (RCTs).⁴⁸ In this design, evaluators (and sometimes implementation contractors) randomly assign sampled members of a population of interest to treatment group or a control group. Among the benefits offered by an RCT—when properly applied—is that it eliminates most selection bias, including free ridership and participant spillover effects. Hence, producing net savings estimates. For some programs, evaluators must take a second step to calculate net savings to ensure savings are not being double-counted, either counting savings being claimed by other programs or savings already credited to earlier program efforts (often called “legacy uplift”).

Free ridership refers to participants in an energy efficiency program that would have saved energy even without the program's stimulus. As these program participants would have engaged in energy-saving actions in the program's absence, counting their savings exaggerates the program's impact. RCTs eliminate free ridership bias because the random assignment of customers to treatment and control groups equally distributes such participants between the two.⁴⁹ Upon comparing the two groups' energy consumption, free ridership energy savings in the control group cancel out those in the treatment group, eliminating free ridership bias.

Participant spillover refers to the tendency of participants in an energy efficiency program to engage in additional energy-saving actions. Though these actions occur outside of the program's scope, they also occur as a direct or indirect result of the program. The extent that these additional savings are not measured and attributed to the program by the evaluator understates the program's impact.

Consideration of participant spillover effects begins by considering what participant spillover means in the context of behavior-based energy efficiency programs. As behavioral programs prescribe neither the installation of any specific measures or sets of measures nor the adoption of any particular behaviors, they likely would not cause participant spillover effects: energy savings resulting from a behavioral program's influence would, by definition,

⁴⁷ State and Local Energy Efficiency Action Network. *Evaluation, Measurement and Verification (EM&V) of Residential Behavior-Based Energy Efficiency Programs: Issues and Recommendations*. Prepared by A. Todd, E. Stuart, S. Schiller, and C. Goldman. Lawrence Berkeley National Laboratory, 2012. (<http://emp.lbl.gov/publications/evaluation-measurement-and-verification-emv-residential-behavior-based-energy-efficiency>)

⁴⁸ For example, most residential, behavior-based energy efficiency programs administered by Opower on behalf of energy utilities are designed as RCTs, as are some commercial and industrial behavioral programs: for example, the EnergyCheck program that Pulse Energy implements for Commonwealth Edison.

⁴⁹ Small differences may occur between the distributions of free ridership's propensity in the two groups for any given sample. Their expected values, however, will be identical, and in any case the size of any such discrepancies shrinks as sample size increases. Thus, this is only a potential concern for programs with unusually small numbers of participants.

be “in scope,” eliminating nearly all possibilities for participant spillover.

The only exceptions would be participant spillover effects not reflected on customers’ bills or meter data. These can arise if spillover savings occur in another venue—outside of the home (e.g., a workplace) for a residential behavioral program—or if the program’s design reduces energy consumption in one form (e.g., electricity) but results in spillover savings in another form (e.g., natural gas). To the extent that either situation occurs, an evaluation relying on an RCT would understate program savings.

In general, RCTs do not address nonparticipant spillover, which reflects a program’s influence on nonparticipants. Such spillover may arise from a behavioral energy efficiency program if, for example, the program indirectly influences customers in the control group or affects the availability of energy efficiency products and services to those served by the relevant market, regardless of whether they participate in the program or belong to the control group. Where significant nonparticipant spillover occurs, an evaluation relying on RCT would understate program savings.

In an RCT, energy consumption of the treatment and control groups can be appropriately compared through a regression analysis, using time-series observations on the usage of individual customers in the treatment and control groups during the pre- and post-treatment periods. Such data most commonly derive from customers’ monthly bill records, hence the frequent use of “billing analysis” to describe this approach (although higher-frequency usage data from customer AMI meters also can be used and provide some additional benefits).⁵⁰ Due to the combined time-series/cross-section structure of such data sets, the NTG Working group recommends that panel regression techniques be used.⁵¹

5.1.2 Non-Randomized Designs

Where randomized assignments prove infeasible, quasi-experimental evaluation methods can be substituted. These methods select a control group using nonrandom methods and are less reliable than RCTs, but, with appropriate care, they can produce valid results. Non-randomized designs can still produce net savings as their primary output, just as RCTs do.

Three quasi-experimental approaches are commonly used to evaluate behavior-based energy efficiency programs that cannot be construed as RCTs:

- Regression discontinuity (RD)
- Variation-in-adoption (VIA)
- Matched controls (MC).

All three create a nonrandom control group to replace a random control group used in the RCT approach.

Regression Discontinuity. RD requires basing a program’s eligibility on a continuous variable (e.g., customers’ adjusted gross income falling below a cutoff value for them to qualify for the program). When this is true, the RD

⁵⁰ These benefits include: having more observations per customer, which improves model precision; obviating concerns over billing periods with differing numbers of days; and providing the ability to observe intraday load shifting in addition to energy savings.

⁵¹ “Panel” refers to the data set consisting of time-series observations on energy consumption of a cross-section of treatment and control customers. Panel estimation techniques refer to the model’s inclusion of terms that control for individual customer heterogeneity (e.g., customer fixed effects or a lagged dependent variable), and cluster-robust standard errors, which can accommodate differing error variances across customers and an intracustomer correlation of errors.

method assumes customers just beyond the cutoff likely will be very similar, on average, to those just inside of it. The method compares changes in energy usage for a group just outside of the eligible range to that of a group of participants just on the other side of the eligibility cutoff. The RD approach, however, is susceptible to an important weakness: misspecification of the regression functional form.⁵²

Variation-in-Adoption. The VIA model applies only to program participants.⁵³ For this method, customers must sign up for the program on a rolling basis. VIA takes advantage of its enrollment's differential timing to compare energy usage of customers opting in to that of customers not yet opting in (but doing so later). The method relies on an assumption that, in any given month, customers have already opted in; those that soon opt in have similar characteristics to those who have enrolled, both in observable and unobservable characteristics. For this assumption to prove valid, customers must decide to opt into the program at different times for essentially random reasons (e.g., influenced only by marketing exposure and program awareness).⁵⁴ In particular, the decision to opt in should not relate to observable or unobservable household characteristics.⁵⁵

Matched Controls. MC creates a control group by matching each treatment customer to the most similar nonparticipant customer available on the basis of exogenous covariates from the pre-enrollment period known to highly correlate with post-enrollment usage.⁵⁶ The covariate most likely to correlate with post-enrollment energy usage in a given time period is customer energy usage during the same period of the preceding year, but other observable factors may be used when available. Implementing MC requires customer usage data for the year preceding all opt-in customers' decisions to participate in the program, along with a large group of nonparticipants who can be assumed to be similar to opt-in customers, aside from their program participation status. The pool of potential matches should be drawn from the same customer class and rate category.

The MC method involves identifying a nonparticipant customer whose energy usage closely matches that of a program participant in the months preceding the participant's enrollment in the program. The logic inherent in this approach is: if the analyst finds a set of nonparticipants who, on average, are the same as participants regarding energy consumption before program enrollment, these matches will provide a good counterfactual estimate of how much energy participants would have used in the program's absence.

The MC approach does present a main weakness: it can only identify matches based on observable customer characteristics, which leaves open the exclusion of the possible influence of relevant unobservables. While factors

⁵² The most common misspecifications are: mistaking a nonlinear relationship for a discontinuity; and failing to recognize potential interactions between assignments and the treatment studied. See W.R. Shadish, T.D. Cook and D.T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, adsworth 2002, pp. 229-238.

⁵³ M. Harding and A. Hsiaw, "Goal Setting and Energy Conservation," July 2013. Available at: http://people.duke.edu/~mch55/resources/Harding_Goals.pdf.

⁵⁴ This differs from an RCT with a recruit-and-delay design, in which customers do not choose when to opt in, but instead are randomly assigned different times to opt in, or from an RCT with a recruit-and-deny design, where customers are randomly denied access to the program.

⁵⁵ As the validity of the VIA method depends on this assumption, it should be empirically tested to the extent possible. If program marketing is punctuated and dates of marketing exposure are known, it is possible to test whether household enrollment in any particular month is driven by marketing activity, as opposed to observed household characteristics or unobserved heterogeneity. A test of whether the energy usage of households before they opt in differs from households that opt in during any particular month as opposed to another month is built into the VIA regression model's functional form. See Harding and Hsiaw, op. cit., for details.

⁵⁶ See Daniel E. Ho, Kosuke Imai, Gary King, and Elizabeth Stuart, 2007, "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15(3): 199-236.

other than pre-enrollment energy usage plausibly could be used (e.g., household income, demographics, geographic location) in the matching process to address relevant unobservable characteristics (e.g., attitudes toward energy conservation and environmental concerns), this assumption cannot be directly tested.⁵⁷

There is a special case of MC called propensity-score matching. This develops a binary choice model to predict the probability that a customer will opt into the program, and then, for a control group, chooses customers with a high propensity for opting in but choosing not to do so. This functions well if observable variables used to calculate the propensity score sufficiently correlate with relevant unobservables to explain differences between treatment and control customers that cannot be explained by matching observables. With most evaluations of energy efficiency programs, however, little (if any) data are available on customers other than their energy usage; so the distinction usually becomes irrelevant.

5.2 Code Compliance Protocol

The protocol represents a basic framework for estimating the NTGR that may be refined based on impact evaluation results. The NTGR is used to convert an estimate of gross savings into an estimate of net savings. Two general methods can be used to estimate gross energy impacts: (1) utility billing data analysis; and (2) building energy modeling.⁵⁸ The specific method used depends on the availability of necessary data.

5.2.1 Data Collection

5.2.1.1 Program Documentation

To inform the NTGR estimate, the evaluator documents program delivery. Information collected includes the following: the number, location, and dates of training workshops; the topics covered; materials disseminated; the number of trainees in each workshop and the type of trainee; and the hours of instruction.

5.2.1.2 Stakeholder Interviews

To inform the NTGR estimate, the evaluator conducts interviews with key stakeholders involved in the program. Interviews should include training program managers, instructors, and trainees. Trainees typically include contractors, builders, consultants, code officials, and others involved in building design and construction. The interviews seek to gather information on how training affected building design, construction, new code compliance, and enforcement.

5.2.2 Attribution Assessment

The NTGR estimation method stays the same, regardless of the method used to estimate gross energy savings.

A Delphi panel⁵⁹ produces an NTGR estimate that reflects the share of gross energy savings resulting from increased code compliance attributable to the program. Formed by selecting four to six knowledgeable

⁵⁷ Such secondary, observable characteristics are rarely available to evaluators of energy efficiency programs, except for geographic location (e.g., postal zone of customer premise).

⁵⁸ The modeled energy savings approach is similar to the approach described by Department of Commerce in Exhibits 6.1 and 6.2 from excerpts of Docket 13-0499 through estimation of potential energy savings.

⁵⁹ The Delphi panel should be conducted according to best practices. For example, see: Day J and Bobeva M (2005) "A Generic Toolkit for the Successful Management of Delphi Studies" *The Electronic Journal of Business Research Methodology* Volume 3 Issue 2, pp. 103-116, available online at www.ejbrm.com.

professionals not associated with the program in any way,⁶⁰ the panel receives estimates of gross energy savings, building construction data, and evidence of attribution—including the results of stakeholder interviews and program documentation. Panel members individually review the information and provide feedback regarding their NTGR estimates and rationales. Responses are compiled, with combined, anonymous responses circulated to all panel members. Panelists review this information, revise their initial estimates and rationales, as they deem appropriate, and provide new estimates and rationales. Evaluators review the second set of estimates and rationales to develop a final attribution estimate, accompanied with a summary of supporting rationales. This NTGR estimate, used in combination with the gross energy savings estimate and building construction data, produces a final estimate of net energy savings attributable to the program.

⁶⁰ Delphi panelists should have no biases that would affect their assessment of the program's effectiveness. Selected individuals should be knowledgeable about building codes and all factors that could conceivably affect code compliance.

6 Appendix A: Overview of NTG Methods

The evaluation teams present information in this appendix to provide a relatively quick overview of NTG methods for readers unaccustomed to the possible methods that evaluators may deploy. It is not meant to be a complete or deep discussion about each of the methods presented. However, the evaluators in Illinois considered the inclusion of this appendix to be very important in acknowledging the current suite of methods deployed by evaluators throughout the U.S. and giving a framework for work within Illinois.

Much of the information shown below is taken directly from a single source—the national Uniform Methods Project, Chapter 23: Estimating Net Savings: Common Practices. (Violette and Rathbun, 2014) This document has done a nice job of summarizing the eight most common attribution methods currently in use across the U.S. The evaluation teams recommend that readers go first to this reference for further information. Additionally, while there are slightly over 100 references within the Violette and Rathbun document, other non-duplicative references are included where reasonable as additional resources for those interested in further research into any specific method.

6.1 Survey-Based Approaches

Virtually all Illinois based evaluations use a survey-based approach for programs where primary data is used to determine net savings. (The main exception is for behavioral programs which use statistical analysis based on a randomized control trial program design.) Survey-based approaches obtain data from program participants and nonparticipants using a structured data collection instrument implemented via phone, in person, or online. At times, evaluators create and use an unstructured depth-interview guide to collect information about attribution, and this provides both contextual data and quantitative data about a given project.

6.1.1 Self-Report Approach

The self-report approach relies on the abilities of customers to discuss the program influence as well as the somewhat abstract ideas of the counterfactual (i.e., what would have occurred absent the program) after making a choice to purchase an energy efficient item or take an energy efficient action unrelated to a purchase. For program participants, this could include doing nothing (i.e., leaving the existing equipment as-is), installing the same energy efficient equipment as they did through the program, or an intermediate step of installing equipment that is more efficient than what they had in place previously, but less efficient than what they installed through the program. Evaluators also use this approach when collecting information from trade allies or distributors. This self-report approach is not new, nor is it exclusively used by the energy efficiency industry. An important attribute of this approach is its reliance on well-designed and fielded survey questions; so that the data underlying subsequent analyses are accurate and complete.

The output of this approach is a NTG ratio which can be considered an index of the program's influence on the decision to install energy-efficient equipment. The NTG ratio is applied to gross savings in order to obtain an estimate of net savings. The NTG ratio may include free ridership, spillover, or market effects, depending on the survey and analytical design. NTG ratios may be calculated at the measure, suite of measures, or program level and are typically average values weighted by savings. If sufficient information is available, analysis of NTG ratios among certain customer segments may be done to further inform changes to program design.

References

- Sudman, 1996
- Stone, et al., 2000
- Bradburn, et al., 2004

6.1.2 Econometric/Revealed Preference Approach

The econometric/revealed preference approach, while still considered a survey approach due to how data is collected, moves beyond asking people about the counterfactual and instead uses the observations of the evaluator to collect information for analysis of a NTG ratio. Within this approach, evaluators typically deploy similar sampling designs as for the self-report approach to collect data, but actively gather what a person is doing (i.e., what is being purchased in a store) to determine attribution.

6.2 Randomized Control Trials and Quasi-Experimental Designs

As mentioned earlier, evaluators deploy randomized control trials (RCT) for estimating savings from the behavioral programs within Illinois. Additionally, quasi-experimental designs (QED) have been used in the past in Illinois to estimate net savings from the upstream CFL program, and CFL, insulation, and air-sealing measures within the Home Performance with ENERGY STAR program.

RCT and QED use statistical analysis to determine regularities within the data that reveal net savings due to a program intervention.⁶¹ The analytical design attempts to control for factors that can confound net analysis.⁶² When estimating net savings within both an RCT and QED, two groups are included within the analysis: 1) a group that has been exposed to (i.e., treated by) a program; and 2) a group that has not been exposed to the program. Evaluators must carefully consider the choice of the non-exposed group (called a control group for RCTs or comparison group for QEDs).

RCT: This design must be integral to a program's implementation. Without the ability to randomly assign customers to one group or another (or at least randomly encourage customers to participate in a program), the ability of the design to yield unambiguous estimates of net impacts is compromised. Evaluators often help design how a program is implemented and, if not involved at the outset, carefully review choices made by the implementation team.

QED: A QED may be designed after a program has been implemented. It relies on determination of an equivalent comparison group, which is often chosen based on energy use. QED is difficult to perform well within the commercial sector due to the heterogeneity of end uses within the sector.

The output of an RCT or QED equals the average net savings for the population within the statistical model. Evaluators may also analyze the data to help understand the savings within specific known segments if sufficient information and data points are available.

References

- Mohr, 1995
- Shadish, Cook, Campbell, 2002
- Scriven, 2008
- Donaldson, 2009

⁶¹ Net savings are calculated when a comparison or control group of non-treated customers are part of the design. Statistical analyses can also obtain gross savings.

⁶² Economists strongly support this approach, but among program evaluators, the idea that an RCT is a "gold standard" for attribution research has been hotly debated for decades.

6.3 Deemed or Stipulated NTG Ratios

A deemed (or stipulated) NTG ratio is a value known prior to implementing a program and applied to estimate net savings for that program in a certain year.

Deemed or stipulated NTG ratios may be based on previous primary data collection, a review of secondary data, or agreed to among stakeholders. In Illinois, deemed or stipulated NTG ratios should reflect best estimates of likely future actual NTG ratios for the relevant program year, taking into consideration stakeholder input, the evaluator's expertise, and the best and most up-to-date information.

6.4 Common Practice Baseline Approaches

For this method, the evaluation team estimates what a typical consumer would have done at the time of the project implementation. Essentially, what is "commonly done" becomes the basis for baseline energy consumption and calculation of net savings. No gross impacts are calculated in this approach. This baseline is defined as the counterfactual "i.e., what would have occurred absent the program" and has been referred to as current practice, common practice, or industry standard practice. Evaluators determine these practices through multiple methods, but often can be from self-report or on-site audits. The difference between the energy use of measures installed in the program and the energy use associated with current practice is considered by some to be sufficiently close to the net savings.

This approach is not in use in Illinois, but it is used elsewhere in the country, such as the Pacific Northwest and Delaware.

6.5 Market Analyses

Market analyses can be done in several ways. Market analyses are often used in theory-driven evaluations of market transformation programs.

Other non-sales data market analyses can be postulated on changes specified in program logic such as: 1) changes in the number of energy-efficient units manufactured; 2) changes in market actor behavior around promotion or stocking of energy-efficient items; or 3) reductions in prices. The analyses involving non-sales data must make a clear link between the program intervention and the changes found in the market. Additionally, outside of Illinois, while evaluators have extrapolated the market changes to specific energy or demand reductions, this activity may be viewed as tenuous due to assumptions that evaluators must make within the analysis.

Illinois is in a position to begin to discuss market analyses and how specific research may be able to interpret changes that have occurred (or may occur in the future) because of the program interventions over the past eight years. Market analyses can be backward looking through historical tracing, but it is best used when the logic of an intervention is described and specific market metrics are tracked over time.

6.6 Structured Expert Judgment Approaches

Closely tied to market analysis, this approach is a way for evaluators to gather credible evidence of changes that arise due to the intervention of a program. When deployed, it is often used as a cost-effective approach to estimate market effects or reach agreement on a NTG value when several different types of evidence are available. The key premise of this approach is the use of a select group of known experts that all stakeholders agree can provide unbiased information as well as having sufficient knowledge to judge what may have occurred absent a program intervention.

A Delphi Panel is an example of this approach where data are collected from two or more rounds of data collection (which can occur via e-mail, Internet, or in person). A round is when experts make their thoughts known about a specific subject; the evaluation team synthesizes the data and provides this collated data back to the group to discuss again. Allowing the full experts to see how their peers think about a topic helps to move the group

towards consensus.

References

- Mosenthal, et al., 2000
- Powell, 2002

6.7 Program Theory-Driven Approach

This approach is not included in the Violette and Rathbun (2014) document as a high-level method, but it is discussed by the authors under the historical tracing method. The Illinois evaluators believe that it deserves at least a short discussion within this framework.

A program theory is the written narrative about why the activities of a program are expected to bring about change. Typically associated with this approach is the direct graphical explication of the linkages between activities, outputs, and outcomes through an impact logic model.⁶³

A theory-driven evaluation denotes “[A]ny evaluation strategy or approach that explicitly integrates and uses stakeholder, social science, some combination of, or other types of theories in conceptualizing, designing, conducting, interpreting, and applying an evaluation.” (Coryn 2011) Within this approach, the ultimate conclusions regarding the efficacy of a program are based on the preponderance of the evidence and not on the results of any single analysis. Coryn and colleagues systematically examined 45 cases of theory-driven evaluations published over a 20-year period to ascertain how closely theory-driven evaluation practices comport with the key tenants of theory-driven evaluation as described and prescribed by prominent theoretical writers. One output from this analysis was the identification of the core principles and sub-principles of theory-driven evaluation. If interested, please review the reference under Coryn 2011.

As an approach, it is best used for complex programs and/or causal mechanisms that extend far into the future. Evaluators collect evidence that supports or rejects hypotheses that are explicit in the logic model. The case for program attribution is strengthened based on the extent to which an evaluation shows that the expected changes occur. Additionally, the evaluation team may be able to collect data that will answer questions about the longer-term outcomes of a program. This type of data collection may be very similar to market tracking activities described briefly above under Market Analyses.

This approach does not specifically estimate a NTG value, but Program Administrators can choose to keep, drop, or change a program based on intermediary data. Regulators must be convinced that the logic of a program is sound and that the intermediary outcomes are causally linked to expected savings.

References

- Weiss, 1997
- Chen, 2000
- Coryn, 2011

6.8 Case Studies Design

Case studies are used extensively in social sciences as well as many other disciplines or practice-oriented areas, such as political science, economics, education, and public policy. Case studies help to understand the how and

⁶³ Evaluators may use logic models to show program processes as well, but this is a program flow chart, not an impact model.

why of a situation and typically retain a holistic aspect of real-life events. As such, they may be a useful approach to determine attribution. As with program theory design, though, the data collected and analyzed within a case study approach will not typically yield a specific NTG value, but can provide credible evidence and insight that supports or refutes the changes brought about by program intervention.

To be used to assess attribution, evaluators must carefully design case studies to assure they account for the threats to causality (i.e., internal validity) that arise in any design. While not typically thought of in this manner, case study design can address multiple types of validity such as construct, internal, and external validity as well as assuring reliability. When establishing construct validity and reliability, evaluators must use multiple sources of evidence, create and maintain a study database, and maintain a “chain of evidence” within the analysis. Internal validity is shown through analytic tactics such as pattern matching, explanation building, addressing rival explanations, or using logic models. External validity centers on the ability to generalize the analytical findings to other similar situations. External validity may be shown through the replication of findings.

References

- Yin, 2003
- Stake, 2006

7 Appendix B: References

- Bradburn, Norman, Sudman, Seymore, and Wansink, Brian. 2004. *Asking Questions. The Definitive Guide to Questionnaire Design – For Market Research, Political Polls, and Social and Health Questionnaires*. San Francisco, CA: Jossey-Bass.
- Chen, Huey-Tsyh. 1990. *Theory-Driven Evaluations*. Sage Publications.
- Coryn, Chris L.S., Norakes, Linday A., Westine, Carl D., Schröter, Daniela C. 2011. "A Systematic Review of Theory-Driven Evaluation Practice from 1990 to 2009." *American Journal of Evaluation*, 32(2): 199-226.
- Donaldson, Steward I., Christie, Christina A., Mark, Melvin M. 2009. *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* Los Angeles, CA: Sage Publications, Inc.
- Mohr, Lawrence B. 1995. *Impact Analysis for Program Evaluation, Second Edition*. Thousand Oaks, CA: Sage Publications, Inc.
- Mosenthal, Philip, Pahl, Ralph, Neme, Chris, and Cuomo, Robert. 2000. *A Modified Delphi Approach to Predict Market Transformation Program Effects*. Proceedings from the 2000 ACEEE Conference.
- Powell, Catherine. 2002. "The Delphi Technique: Myths and Realities." *Journal of Advanced Nursing*. 41(4), 376-382.
- Scriven, Michael. "A Summative Evaluation of RCT Methodology and an Alternative Approach to Causal Research." March 2008. *Journal of MultiDisciplinary Evaluation*, Volume 5, Number 9. ISSN 1556-8180.
- Shadish, William R., Cook, Thomas D., and Campbell, Donald T. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin Company.
- Stake, Robert E. 2006. *Multiple Case Study Analysis*. The Guilford Press.
- Stone, Arthur A., Turkkan, Jaylan S., Bachrach, Christine A., Jobe, Jared B., Kurtzman, Howard S., Cain, Virginia S. editors. 2000. *The Science of Self-Report. Implications for Research and Practice*. Lawrence Erlbaum Associates, Inc.
- Sudman, Seymour, Bradburn, Norman, Schwarz, Norbert. 1996. *Thinking about Answers. The Application of Cognitive Processes to Survey Methodology*. Jossey-Bass.
- Violette, Daniel M., Rathbun, Pamela. 2014. "Chapter 23: Estimating Net Savings: Common Practices. The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures." Available electronically at <http://www.nrel.gov/docs/fy14osti/62678.pdf>.
- Weiss, Carol H. 1998. *Evaluation Methods for Studying Programs and Policies, 2nd Edition*. Upper Saddle River, New Jersey: Prentice Hall.
- Yin, Robert K. 2003. *Case Study Research. Design and Methods*. Thousand Oaks, CA: Sage Publications, Inc.